2019

Cell-To-Cell Heterogeneity In Transcription And Splicing

Allison Cote
University of Pennsylvania

Cell-To-Cell Heterogeneity In Transcription And Splicing

Abstract

Biological variability is often treated as a drawback of biological systems, but we have used it to learn about various biological phenomena, such as mesenchymal stem cell variability and the timing of splicing. We quantified the variability of mesenchymal stem cell clones with respect to their ability to generate cartilage-like extracellular matrix tissues and used this variability to identify markers that could be used to sort cells that produce high quality cartilage tissue. This revealed that canonical markers commonly used in cartilage tissue engineering are poor markers of high performing mesenchymal stem cells, suggesting that sorting cells based on these markers will not enrich populations for cartilage tissue production. We also used RNA FISH to characterize single-cell level expression of these markers and found that the expression showed high variability and very little cell-to-cell memory. RNA FISH also revealed that during de-differentiation of chondrocytes in monolayer culture, expression of canonical markers does not decrease on the absolute level, but only relative to cell size, also suggesting the gene expression level of these markers is not the sole determining factor for the production of high-quality cartilage tissue.

I then decided to use RNA FISH to quantify the gene-to-gene and intron-to-intron heterogeneity in splicing that occurs at and near the site of transcription, revealing a spectrum of co- and post-transcriptional splicing of endogenous genes. Observing multiple introns of the same gene simultaneously showed that introns are largely spliced independently of one another and not in a 5' to 3' order. Using a combination of RNA FISH and expansion microscopy, I show that transcripts dwell at the transcription site after transcription is complete, suggesting that fractionation methods may underestimate "post-transcriptionality". Taken together, our observations suggest a model of dwell time at the site of transcription, variable intron retention, and a sequence-specific regulation of the timing and localization of splicing for individual introns.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Cell & Molecular Biology

First Advisor

Arjun Raj

Keywords

cartilage tissue engineering, RNA FISH, splicing, transcription

Subject Categories

Biology | Cell Biology | Molecular Biology

CELL-TO-CELL HETEROGENEITY IN TRANSCRIPTION AND SPLICING

Allison Jean Cote

A DISSERTATION

in

Cellular and Molecular Biology

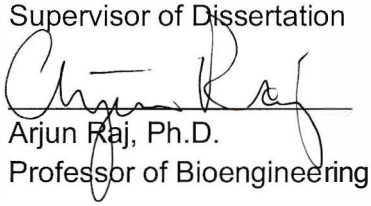
Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy

2020

Supervisor of Dissertation



Arjun Raj, Ph.D.
Professor of Bioengineering and Genetics

Graduate Group Chairperson

Daniel S. Kessler, Ph.D.
Associate Professor of Cell and Developmental Biology

Dissertation Committee:
Jennifer E. Phillips-Cremins, Ph.D.
Assistant Professor of Bioengineering
Robert L. Mauck, Ph.D.
Mary Black Ralston Professor of Orthopaedic Surgery, Professor of Bioengineering
Rebecca G. Wells, M.D.
Professor of Medicine, Bioengineering, and Pathology & Laboratory Medicine
Christopher J. Lengner, Ph.D.
Associate Professor of Biomedical Sciences

Acknowledgements

I would like to thank my parents, for their endless love and support during my many years as a graduate student.

I would like to thank John for his help and support over the past year, I would not have gotten through this without him.

I discovered during my time as a graduate student that I have a vast support network and would like to thank all of my friends and family for supporting me through the many tough years.

Lastly, I would like to thank Arjun, who never gave up on me. You have been an invaluable source of support and scientific guidance and will have a lasting impact on how I think and communicate with others.

ABSTRACT

CELL-TO-CELL HETEROGENEITY IN TRANSCRIPTION AND SPLICING

Allison Jean Coté

Arjun Raj, Ph.D.

Biological variability is often treated as a drawback of biological systems, but we have used it to learn about various biological phenomena, such as mesenchymal stem cell variability and the timing of splicing. We quantified the variability of mesenchymal stem cell clones with respect to their ability to generate cartilage-like extracellular matrix tissues and used this variability to identify markers that could be used to sort cells that produce high quality cartilage tissue. This revealed that canonical markers commonly used in cartilage tissue engineering are poor markers of high performing mesenchymal stem cells, suggesting that sorting cells based on these markers will not enrich populations for cartilage tissue production. We also used RNA FISH to characterize single-cell level expression of these markers and found that the expression showed high variability and very little cell-to-cell memory. RNA FISH also revealed that during de-differentiation of chondrocytes in monolayer culture, expression of canonical markers does not decrease on the absolute level, but only relative to cell size, also suggesting the gene expression level of these markers is not the sole determining factor for the production of high-quality cartilage tissue.

I then decided to use RNA FISH to quantify the gene-to-gene and intron-to-intron heterogeneity in splicing that occurs at and near the site of transcription, revealing a spectrum of co- and post-transcriptional splicing of endogenous genes. Observing multiple introns of the same gene simultaneously showed that introns are largely spliced

independently of one another and not in a 5' to 3' order. Using a combination of RNA FISH and expansion microscopy, I show that transcripts dwell at the transcription site after transcription is complete, suggesting that fractionation methods may underestimate "post-transcriptionality". Taken together, our observations suggest a model of dwell time at the site of transcription, variable intron retention, and a sequence-specific regulation of the timing and localization of splicing for individual introns.

List of Tables

List of Figures

Chapter 2: Matrix expression does not predict matrix deposition

Chapter 3: Variability in splicing from one intron to the next

Chapter 1: Introduction

Why are some cells different from others? Cells can differ from one another in many ways. A liver cell, for example, is different in many ways from a red blood cell. Even in the case of bacteria, there is variability from one cell to the next. This variability can be expressed through protein, RNA, DNA, sugars, lipids, ions, water, or any other type of molecule in the cell. Each type of variability has different consequences for cell function. Between cell types there are often obvious differences in cellular function, but these differences are frequently overlooked within seemingly identical cell populations. Many bulk assays obscure these differences by taking averages across thousands or millions of cells. Traditional gene expression assays, like microarrays or bulk RNA-sequencing, take averages across thousands of cells, even though there is marked heterogeneity in gene expression at the single cell level (Cote et al., 2016; Maamar et al., 2007; Raj et al., 2008, 2006; Raj and van Oudenaarden, 2009; Tang et al., 2009).

Gene expression can be regulated at many levels, from structural organization of the DNA to post-translational modifications of proteins. With the advent of high throughput sequencing technologies much focus has been placed on gene expression at the RNA level. Sequencing and imaging are now routinely applied at the single-cell level to characterize gene expression (Femino et al., 1998; Hwang et al., 2018; Raj et al., 2006; Stuart and Satija, 2019; Tang et al., 2009). Sequencing has been used to quantify many aspects of RNA expression variability, from the transcription of nascent RNA to RNA half-life and degradation (Lugowski et al., 2018; Mayer et al., 2015). These sequencing methods can also yield information about intermediates of RNA processing, such as

splicing, capping, polyadenylation, editing, and modifications (Gruber et al., 2016; Motorin and Helm, 2019; Ramaswami et al., 2013; Takahashi et al., 2012; Waks et al., 2011).

Other techniques, such as mass spectrometry to look at protein level heterogeneity, are still nascent in their application to single-cells (Budnik et al., 2018), but are areas of fervent development. All single cell techniques have revealed cell-to-cell heterogeneity in their respective measurements. Traditionally, it was assumed that cell-to-cell differences were due to genetic differences or differences in environment. However, single cell sequencing and imaging has revealed a variety of gene expression states at the RNA level, even within seemingly identical populations of cells (Symmons and Raj, 2016). This heterogeneity makes results from bulk assays hard to interpret and suggests that there is some underlying reason why genetically identical and environmentally similar cells have differences in gene expression.

Many researchers, in a variety of contexts, have suggested that this heterogeneity in seemingly identical cells could be due to transient changes in gene expression (for a review, see (Raj and van Oudenaarden, 2008; Symmons and Raj, 2016)). These transient changes in gene expression have resulted in distinct phenotypes for these cells, such as drug resistance (Shaffer et al., 2017). However, this phenotype (drug resistance) was only possible with the transition of this transient state to a more permanent, heritable state of gene expression (Shaffer et al., 2017). Heritability, or the ability of a cell to “remember” its molecular state and recapitulate a similar state after the disruptive process of cell division, is especially important to consider in cases where cells can be separated by their gene expression state and then left to grow over time. In a case where gene expression was not heritable, cells isolated based on their expression of a particular gene (which is an extremely frequent method for isolating various cell types) could, soon after isolation,

express a completely different level of that gene. This lack of heritability could be one explanation for the high level of variability within a population of cells that was seemingly identical at a point in time.

In this thesis, I have measured this heterogeneity in gene expression at two distinct regulatory steps in two different systems:

1) Transcriptional heterogeneity in mesenchymal stem cells and chondrocytes
2) Splicing heterogeneity in cancer cells

I used this cell-to-cell heterogeneity in transcription and splicing to answer fundamental questions about the memory of transcription, predicting mesenchymal stem cell function, and the timing and spatial distributions of splicing.

Heterogeneity of mesenchymal stem cells makes cartilage tissue engineered constructs inconsistent

Cartilage degradation affects millions of people each year, causing a great deal of pain and loss of mobility (Chubinskaya et al., 2015). This degradation can be the result of many factors, such as age, BMI, joint stress, osteoarthritis, etc., but one universal result is an inability of cartilage to repair itself. Articular cartilage (a smooth, white cartilage found in joints) is frequently damaged by injury or by repetitive stresses from normal motions, such as walking (Chubinskaya et al., 2015; Coleman et al., 2017). Several joints, particularly those in the knees, are under constant mechanical stresses even within the realm of normal motion. These mechanical stresses can damage and wear away the cartilage over time, forcing the loss of lubrication of joints that cartilage provides.

It is difficult for cartilage to repair itself for two primary reasons: 1) cartilage is highly acellular, primarily consisting of extracellular matrix and 2) cartilage is highly avascular,

which prevents access to the necessary nutrients that might otherwise be needed for repair (Carballo et al., 2017; Sophia Fox et al., 2009; Woo and Buckwalter, 1988). The lack of cells in cartilage means that, once some cartilage is worn away, there are often very few cells nearby to replenish the extracellular matrix. For these reasons, researchers have turned to cartilage tissue engineering for repair of cartilage defects.

The ultimate goal of cartilage tissue engineering is to create cartilage that can replace degraded cartilage tissue *in vivo*. To that end, researchers have turned to a combination of cells, scaffolding, soluble cues (such as cytokine treatment), and mechanical stresses to recreate cartilage *in vitro* (Huang et al., 2016; Rai et al., 2017). Each one of these factors can be manipulated individually or in combination to achieve cartilage like substances that are candidates for implantation to repair cartilage tissue, but it remains difficult to successfully reproduce cartilage substitutes with the same composition and mechanical qualities of native human cartilage .

One of the many difficulties in reproducing cartilage is derived from the fact that cartilage, unlike many other tissues, is highly acellular, with chondrocytes (the primary - often considered only - cell type in cartilage) comprising only ~2% of cartilage volume (Alford and Cole, 2005). Chondrocytes play an important role in local repair of cartilage substrate, but they are trapped by the vast extracellular matrix, comprising the majority of cartilage tissue volume (Sophia Fox et al., 2009). This extracellular matrix is comprised of primarily water (~80% of tissue volume) but this water is contained by a series of structural proteins (primarily type II collagen) and other macromolecules, primarily proteoglycans (Maroudas et al., 1991). Briefly, collagen is a structural protein that allows articular cartilage to resist tensile stresses and proteoglycans are protein molecules bound to glycosaminoglycans (sugars), which allow articular cartilage to resist compressive stresses. It is these

extracellular matrix components that are frequently measured to determine success at *in vitro* cartilage formation, and are therefore traditionally used as chondrocyte "markers" (Mackay et al., 1998; Pittenger et al., 1999).

Researchers have naturally turned to cartilage development to attempt to determine how to produce cartilage similar to *in vivo* human cartilage. The earliest stages of cartilage development involve condensation of mesenchymal cells and transient upregulation and deposition of certain extracellular matrix genes (DeLise et al., 2000). Then cells begin depositing specific extracellular matrix components that make up the final cartilage extracellular matrix (DeLise et al., 2000). Researchers have attempted to mimic this cell condensation with pellet based mesenchymal stem cell culture (Bosnakovski et al., 2006). In addition, cues of all types have been taken from cartilage development, such as a soft (compared to bone) substrate for cells (similar to that found in the cell condensation stage of cartilage development), TGF-β signaling, and mechanical stresses (DeLise et al., 2000; Mauck et al., 2006, 2000; Miura et al., 1994).

Many substrates have been used to seed cells for cartilage tissue engineering, from hydrogels to actual decellularized cartilage matrix (Drury and Mooney, 2003; Rai et al., 2017; Schwarz et al., 2012). These matrices span dimensions, from 2D to 3D, and even matrices that change over time (Liao et al., 2015; Panadero et al., 2016; Yoshimoto et al., 2003). Hydrogels are popular scaffolds for cartilage tissue engineering (Drury and Mooney, 2003) because they can mimic the natural composition of cartilage, including collagen type II and glycosaminoglycans. One popular substrate for cartilage tissue engineering is agarose, a relatively soft hydrogel that supports chondrogenic matrix deposition while encapsulating the cells in a physiologically 3D environment (Benya and Shaffer, 1982; Mauck et al., 2006, 2000).

In addition to various physical cues that the substrate provides, cartilage tissue engineering also frequently employs soluble cues to encourage cartilage differentiation of the cells within the matrix and deposition of additional extracellular matrix components (Puetzer et al., 2010). One common soluble cue given to cartilage tissue engineering constructs is TGF-β, a cytokine implicated in cartilage tissue development (DeLise et al., 2000; Miura et al., 1994; Puetzer et al., 2010; Wang et al., 2014). TGF-β is implicated in all stages of cartilage creation, from the initial mesenchymal cell condensation to maintenance of adult articular cartilage (Wang et al., 2014).

TGF-β and a soft 3D matrix can both combine to encourage deposition of extracellular matrix components essential for cartilage function, such as collagen type II and glycosaminoglycans. However, the choice of cells to receive these soluble and physical cues remains of utmost importance. Many cell types have been used in cartilage tissue engineering, such as adult and juvenile chondrocytes, mesenchymal stem cells (the cell type that differentiates into chondrocytes during development), embryonic stem cells, and induced pluripotent stem cells (Benya and Shaffer, 1982; Castro-Viñuelas et al., 2018; Colter et al., 2001; Huang et al., 2010; Toh et al., 2011; Toh and Cao, 2014). However, each cell type comes with drawbacks. For example, chondrocytes have limited proliferative capacity and often begin to lose expression of essential genes over time in culture after their initial isolation (Darling and Athanasiou, 2005; Lin et al., 2008; Ma et al., 2013). Mesenchymal stem cells, as well as embryonic and induced pluripotent stem cells, all need to be differentiated into a chondrocyte like cell before they begin depositing cartilage matrix proteins. However, mesenchymal stem cells are much further along the differentiation path to chondrocytes than either embryonic or induced pluripotent stem

cells, and therefore frequently exhibit greater matrix deposition in response to similar stimuli (Beane and Darling, 2012).

However, even within the population termed “mesenchymal stem cells” there are many sources of heterogeneity. Mesenchymal stem cells are defined as cells that stick to plastic, express a particular set of markers (CD105, CD73 and CD90, and lack CD45, CD34, CD14 or CD11b, CD79alpha or CD19 and HLA-DR), and have the capacity to differentiate into osteoblasts, adipocytes and chondroblasts (Dominici et al., 2006). However, even within these exclusively selected cells there is heterogeneity (McLeod and Mauck, 2017). Mesenchymal stem cells can be isolated from many tissue locations, including bone marrow, adipose tissue, dental tissue, and many others, and each isolation location can yield mesenchymal stem cells of varying proliferative and matrix deposition capacity (Ng et al., 2016). Mesenchymal stem cells can also be isolated from a variety of different species, and each results in cartilage tissue constructs with varying properties (Sakata et al., 2015). Mouse cartilage, for example, is much more densely cellular than human cartilage (Stockwell, 1971), and therefore the properties of mesenchymal stem cells that form this cartilage are likely to be quite different. In a direct comparison of rat and human bone marrow derived mesenchymal stem cells, it was shown that rat mesenchymal stem cells can exhibit increased cartilage like matrix deposition as compared to human mesenchymal stem cells (Scuteri et al., 2014), suggesting that rat and mouse cells may not accurately model human cell behavior in this context. For this reason, many researchers have turned to isolating mesenchymal stem cells from larger animals with cartilage and cartilage defects more similar to human cartilage (Stockwell, 1971), such as cows (Mauck et al., 2006).

In addition to variation due to the location of isolation or cell species, the context of the specific donor of mesenchymal stem cells also matters (Phinney et al., 1999), especially in terms of donor age (D'Ippolito et al., 1999). However, even clonal populations (derived from single cells and therefore necessarily of the same isolation location, species, and donor) of mesenchymal stem cells exhibit varying capacity to differentiate into chondrocytes and other lineages (Muraglia et al., 2000; Russell et al., 2010). This suggests that there are necessarily non-genetic mechanisms that contribute to mesenchymal stem cell differentiation heterogeneity, and sequencing of naïve mesenchymal stem cells has shown that even without differentiation cues, clonal mesenchymal stem cells display heterogeneity in their transcriptomes (Delorme et al., 2009; Freeman et al., 2015). This heterogeneity is apparently especially with regards to markers that may suggest differentiation of cells towards a particular phenotype, or lineage "priming" (Delorme et al., 2009; Freeman et al., 2015). These markers used to define phenotype or priming are often the markers of cartilage tissue itself, such as collagen type II (Delorme et al., 2009) or the transcription factors directly upstream of the markers of cartilage tissue differentiation, such as SOX9 (Freeman et al., 2015). These markers display marked heterogeneity in both the naive and differentiated mesenchymal stem cell populations (Cote et al., 2016; Delorme et al., 2009; Freeman et al., 2015). That begs the question, what does being a "marker" or a specific subpopulation really mean? Does that mean constant expression in that particular cell type and no expression in other cell types? If so, this suggests that traditional extracellular matrix associated gene expression is a poor marker of chondrogenic mesenchymal stem cells.

In this thesis, I measured the cell-to-cell variability in gene expression of mesenchymal stem cells and chondrocytes using RNA FISH and RNA sequencing. RNA FISH is a

technique that allows for visualization of individual mRNA in single cells (Femino et al., 1998; Raj et al., 2006), and therefore let me identify gene expression signatures that suggested lineage priming in specific subsets of mesenchymal stem cells. I then used this lineage priming phenomenon to attempt to identify mesenchymal stem cells that would ultimately form the most cartilage extracellular matrix. I showed that canonical cartilage tissue markers are NOT good predictors of ultimate mesenchymal stem cell function, and that absolute RNA expression of these markers were also not a good indicator of chondrocyte dedifferentiation. A combination of RNA FISH and timelapse imaging showed that there is very little transcriptional memory of canonical cartilage tissue markers in mesenchymal stem cells, which may explain why they did not function as good *a priori* identifiers of high functioning mesenchymal stem cells.

Fractionation and imaging yield conflicting information about the timing of splicing

The life of an RNA begins at transcription and ends at degradation, with many potential regulatory steps in between (such as capping, polyadenylation, modifications, editing, splicing, nuclear export, translation, etc.). Each step has been extensively characterized using high throughput sequencing, among other methods (Gruber et al., 2016; Ingolia, 2016; Motorin and Helm, 2019; Ramaswami et al., 2013; Stewart, 2019; Takahashi et al., 2012; Valencia et al., 2008). One step in particular, splicing, has been shown to not only be an essential process for the majority of eukaryotic genes (Faustino and Cooper, 2003), but the discovery of alternative splicing has made it clear that splicing is also essential for creating the proteasomal diversity that is required by complex human development and function (Wang et al., 2015).

Splicing is the process by which precursor messenger RNA (pre-mRNA) becomes mature mRNA, via the excision of long, non-protein-coding regions of RNA (introns) from the rest of the RNA that will ultimately be translated into protein (exons). Splicing is an essential process in all eukaryotes and defects in both constitutive and alternative splicing cause a wide array of human diseases (Faustino and Cooper, 2003). Splicing is not only responsible for the removal of introns, but is also associated with various other RNA processing and quality control steps, such as nuclear export (Valencia et al., 2008). Alternative splicing is also important for eukaryotic development (Baralle and Giudice, 2017), and over 95% of genes have been shown to be alternatively spliced in human cells (Pan et al., 2008).

Though splicing has been shown to be an essential process for eukaryotic life (Faustino and Cooper, 2003), it is still unclear exactly when splicing happens. When the timing of splicing is altered, for example by the S34F mutation in the gene U2AF1, a core member of the spliceosomal machinery, it can result in alternative splicing and a drastic delay in the timing of splicing (Coulon et al., 2014). The S34F mutation in U2AF1 is frequently found in a variety of cancers as well (Kim et al., 2018; Palangat et al., 2019). Different types of measurements and measurements from different genes yield results that vary several orders of magnitude (Alpert et al., 2016). Different species also have drastically different gene structures, suggesting that conclusions from one may not be applicable to another (Keren et al., 2010). Yeast, for example, have short introns that are thought to be spliced out almost instantly (within a few seconds) after the process of transcription is completed (Wallace and Beggs, 2017). In species with much longer introns, such as humans and mice, splicing has been shown to potentially take much longer, with measurements ranging from 15 seconds up to 5-10 minutes (Alpert et al., 2016).

Generally, the timing of splicing is broken down into two categories that compare splicing rate with the rates of other processing steps. These categories are “co-transcriptional” or tracking along with the polymerase as it transcribes, and “post-transcriptional”, or taking place after the process of transcription (and often polyadenylation as a surrogate for the end of transcription) has been completed. The dominant view of the field is that splicing happens primarily co-transcriptionally (Ameur et al., 2011; Bentley, 2014; Brugiolo et al., 2013; Custódio and Carmo-Fonseca, 2016; Davis-Turak et al., 2015; Gazzoli et al., 2016; Goldstrohm et al., 2001; Herzel et al., 2017; Morlando et al., 2008; Nojima et al., 2018; Pandya-Jones and Black, 2009; Saldi et al., 2016; Singh and Padgett, 2009; Tilgner et al., 2012). However, for many reporters and endogenous genes it has been shown that splicing is happening post-transcriptionally (Coulon et al., 2014; Vargas et al., 2011; Waks et al., 2011). Changing the timing of splicing by changing the elongation rate of RNA polymerase can cause drastic changes in alternative splicing *in vivo* (de la Mata et al., 2003). Changing the timing of splicing via the S34F mutation in splicing factor U2AF1 can make splicing more “post-transcriptional” (Coulon et al., 2014).

The timing of splicing has been shown to be important in many cases, but especially in cases of alternative splicing of certain genes (Waks et al., 2011). In cases where an RNA has many alternate isoforms, it could be necessary for the cell to completely synthesize the RNA before making its splicing “choice”. Thus, any splicing in that case would be completely post-transcriptional.

The debate about the timing of splicing has been occurring for decades (Beyer et al., 1981; Tsai et al., 1980). Some of the first evidence for co-transcriptional splicing came from EM images of Drosophila chromosomes spreads, which revealed nascent RNA still attached to chromatin being spliced (Beyer et al., 1981). The first evidence for post-

transcriptional splicing came around the same time, when Tsai *et al.* observed high molecular weight poly-adenylated transcripts, which they suggested were unspliced, but fully transcribed, intermediates (Tsai et al., 1980).

Currently, there are many methods by which researchers attempt to pinpoint the timing of splicing. Current imaging based methods include cyro-EM, live cell imaging of fluorescent reporter constructs, and RNA FISH of intronic and exonic sequences simultaneously (Coulon et al., 2014; Martin et al., 2013; Vargas et al., 2011; Waks et al., 2011; Wilkinson et al., 2018). These methods all have the potential to reveal different information about the timing of splicing. Cyro-EM has been primarily used to identify the structures, and changes to those structures, that are essential for the splicing process (Wilkinson et al., 2018). Live cell imaging has been used to directly visualize the timing of splicing occurring in single, living cells, but this has primarily been for reporter constructs and modified endogenous genes (Coulon et al., 2014; Martin et al., 2013). RNA FISH has been used to show that splicing primarily happens at the site of transcription, but alternative splicing or insertion of polypyrimidine tract can spatially decouple transcription and splicing (Vargas et al., 2011; Waks et al., 2011; Zhang et al., 1994). Though imaging has yielded a great deal of information about the structure and timing of splicing intermediates, the primary ways by which the timing of splicing are characterized are fractionation and sequencing (or qPCR) based methods (Alexander et al., 2010a, 2010b; Carrillo Oesterreich et al., 2016, 2010; Churchman and Weissman, 2011; Harlen et al., 2016; Milligan et al., 2016; Pandya-Jones and Black, 2009).

There are many current fractionation based methods that have classified splicing as “co-transcriptional” based on measuring the splicing status of RNA after different methods of isolating nascent RNAs. Chromatin fractionation is a popular method of isolating nascent

RNAs, whereby cells are separated into cytoplasmic, nucleoplasmic, and chromatin fractions (Tilgner et al., 2012). This separation is traditionally achieved by a series of increasingly harsh detergent concentrations, which lyse the cytoplasm first, allowing nuclear isolation, then lyse the nucleus and isolate chromatin by sedimentation via centrifugation (Mayer et al., 2015). Nojima *et al.* and others isolate nascent RNA by immunoprecipitating polymerase II and sequencing the RNA that is directly associated with various post-translational modifications of polymerase II (Harlen et al., 2016; Harlen and Churchman, 2017; Nojima et al., 2015). Nascent RNA can also be purified through various metabolic labelling techniques, in which a isolatable nucleotide analogue is incorporated into RNA being produced during a short period of time, and then RNA species with the metabolic label are isolated and their splicing status is measured (Barrass et al., 2015; Eser et al., 2016; Rabani et al., 2014, 2011; Windhager et al., 2012). All of these methods have suggested, thus far, that splicing largely happens immediately after the process of transcription, based on the presence of spliced RNA in the "nascent RNA" fractions. However, new methods that combine both metabolic labelling and chromatin fractionation are beginning to suggest that the presence of spliced RNA in the nascent RNA fractions may represent contamination of already transcribed, fully processed mRNA, and that, when the two methods of 4sU and chromatin isolation are combined, it reveals a much more "post-transcriptional" view of human splicing (Drexler et al., 2019).

Though sequencing methods are now frequently applied at the single-cell level, information about the timing of splicing at the single-cell level is rare, with the notable exception of imaging based methods (Coulon et al., 2014; Martin et al., 2013; Vargas et al., 2011; Waks et al., 2011). However, it is clear that alternative splicing plays a large role in development, even at the single cell level (Lukacsovich et al., 2019; Tang et al., 2009),

and single cell sequencing methods have revealed a variety of splicing isoforms in single cells (Tang et al., 2009).

In this thesis, I will characterize the single-cell level spatial distributions of splicing in various cancer cell lines using RNA FISH. I use this information to infer information about the timing of splicing within these individual cells, including showing that at least some splicing of each observed gene is happening post-transcriptionally. I also employ a combination of RNA FISH and expansion microscopy (Chen et al., 2016, 2015) to visualize individual RNA at the site of transcription and use this to infer that transcripts are not tethered to the site of transcription, and that transcripts dwell at the site of transcription after the process of transcription is completed, complicating results from fractionation assays. Taken together, my data suggests a model of variable timing of splicing of individual introns and a dwell time of transcripts at the site of transcription.

Chapter 2: Matrix expression does not predict matrix deposition

Introduction

Regenerative medicine strategies such as tissue engineering combine advances in cell biology, biomaterials and medicine to restore tissue function. Some approaches utilize stem cells for regeneration. For example, researchers commonly use multipotent progenitor cells, including mesenchymal stem cells (MSCs), for tissue engineering due to their capacity to undergo either osteogenic, adipogenic or chondrogenic differentiation (Pittenger et al., 1999). However, even with the most effective differentiation protocols, individual MSCs demonstrate heterogeneity in their biophysical properties and in their ability to undergo lineage commitment(Colter et al., 2001; González-Cruz et al., 2012; Mareddy et al., 2007; Phinney, 2012), with some clonal subpopulations robustly committing to a differentiated fate while other clones fail to respond to differentiation cues (Muraglia et al., 2000; Phinney, 2012; Russell et al., 2010). Furthermore, in cases in which it seems as though all cells have differentiated based on bulk expression of a particular marker, individual cells within the population may continue to express markers of other lineages (Ponce et al., 2008; Song, 2004). Given that underperforming, alternatively performing or non-responsive subpopulations will hinder the performance of engineered tissues, this inherent MSC heterogeneity compromises therapeutic efficacy. As such, quantitative strategies to select 'superior' subpopulations a priori would improve translational potential.

Despite the phenotypic heterogeneity in MSC populations, most papers that explore the molecular underpinnings of phenotype monitor differentiation via bulk assays of transcriptional state and protein synthesis averaged over an entire cell population. These ensemble measurements, by definition, mask population heterogeneity (Altschuler and

Wu, 2010; Nimmo et al., 2015). The advent of single-cell methods allows for the measurement of cell-to-cell variation and the ability to quantify absolute gene expression in a single cell (Crosetto et al., 2015; Itzkovitz and van Oudenaarden, 2011; Junker and van Oudenaarden, 2014), revealing, for example, marked transcriptional heterogeneity. Real-time fluorescent monitoring of changes in transcript levels in individual cells has also shown that individual MSCs differ in the timing and extent to which they upregulate an early osteogenic marker (Marble et al., 2014). These findings underscore the limitations of coarse ensemble approaches and highlight the need for single-cell molecular profiling of these differentiation events. Although it is reasonable to speculate that the subpopulation of cells expressing high levels of marker genes would ultimately be the most chondrogenic, this hypothesis remains untested.

Given that individual MSCs are highly variable in their capacity to undergo chondrogenesis and accumulate cartilage-like matrix (Huang et al., 2010), we postulated that one could use single-cell marker gene transcript levels as a means to enrich for MSC subpopulations most suited for therapeutic application. Here we define this relationship by developing probe sets for RNA fluorescence in situ hybridization (FISH) directed against transcripts of markers of cartilage, bone and fat, and use single-cell analysis to delineate the relationships between absolute transcript level and differentiated cell function. Specifically, we hypothesized that cells that robustly accumulate an aggrecan-rich, cartilage-like matrix would also express high levels of aggrecan mRNA, while at the same time suppressing markers of other lineages.

We find surprising levels of variability in the expression of aggrecan and other marker genes between individual MSCs both before and after differentiation. However, when we compare the expression with functional capacity (defined by actual matrix deposition) on

a single-cell basis, we find a weak correlation between transcript abundance and protein expression. Transcriptome-wide analysis via RNA sequencing further suggests that neither an expanded set of marker genes, nor the principal components of global gene expression variation, correlate strongly with functional capacity. Indeed, even in fully differentiated chondrocytes derived from native tissue, absolute aggrecan mRNA expression is decoupled from cartilage-like matrix accumulation. Collectively, these findings suggest that sorting based solely on a small set of differentiation markers will not improve chondrogenic outcomes, and challenge the traditional notion that marker gene expression defines or is even strongly associated with phenotype.

Results

Single cells express differentiation markers heterogeneously.

To quantify absolute gene expression of marker genes on a single-cell basis during MSC differentiation and chondrocyte de-/re-differentiation, we paired classic cartilage tissue engineering and cartilage biology experiments with single-molecule RNA FISH (Femino et al., 1998; Raj et al., 2008). Specifically, we monitored the simultaneous expression of aggrecan as a marker of chondrogenic differentiation, GAPDH as a reference gene, and osteopontin and lipoprotein lipase (LPL) as markers of alternate fates (osteogenesis and adipogenesis, respectively) (Frank et al., 2002; Pelttari et al., 2008; Sekiya et al., 2003). For each gene, we designed fluorescently labelled sequence-specific oligonucleotide probes to visualize individual mRNA molecules in intact fixed cells. Individual mRNA appeared as bright diffraction limited spots (Fig. 1a,b), and subsequent spot counting yielded absolute copy number at the single-cell level.

To show that our measurements corresponded well with existing measurements of these systems, we first determined how absolute gene expression changed as MSCs underwent chondrogenic differentiation. To do so, we formed engineered constructs and used RNA FISH to quantify gene expression over 3 weeks in chemically defined media with or without transforming growth factor-b (TGFb; chondrogenic induction media and control media, respectively, Fig. 1c). As expected, chondrogenic induction promoted proteoglycan synthesis and matrix accumulation (Fig. 1d) and increased aggrecan copy number in comparison with control media (Fig. 1e). Although there was considerable donor-to-donor variability in mean aggrecan levels and matrix deposition, the trends were similar between donors, with mean aggrecan copy number generally increasing over the first 7 days, before decreasing at later time points (Fig. 1e, Supplementary Fig. 1A and Table 1). Mean

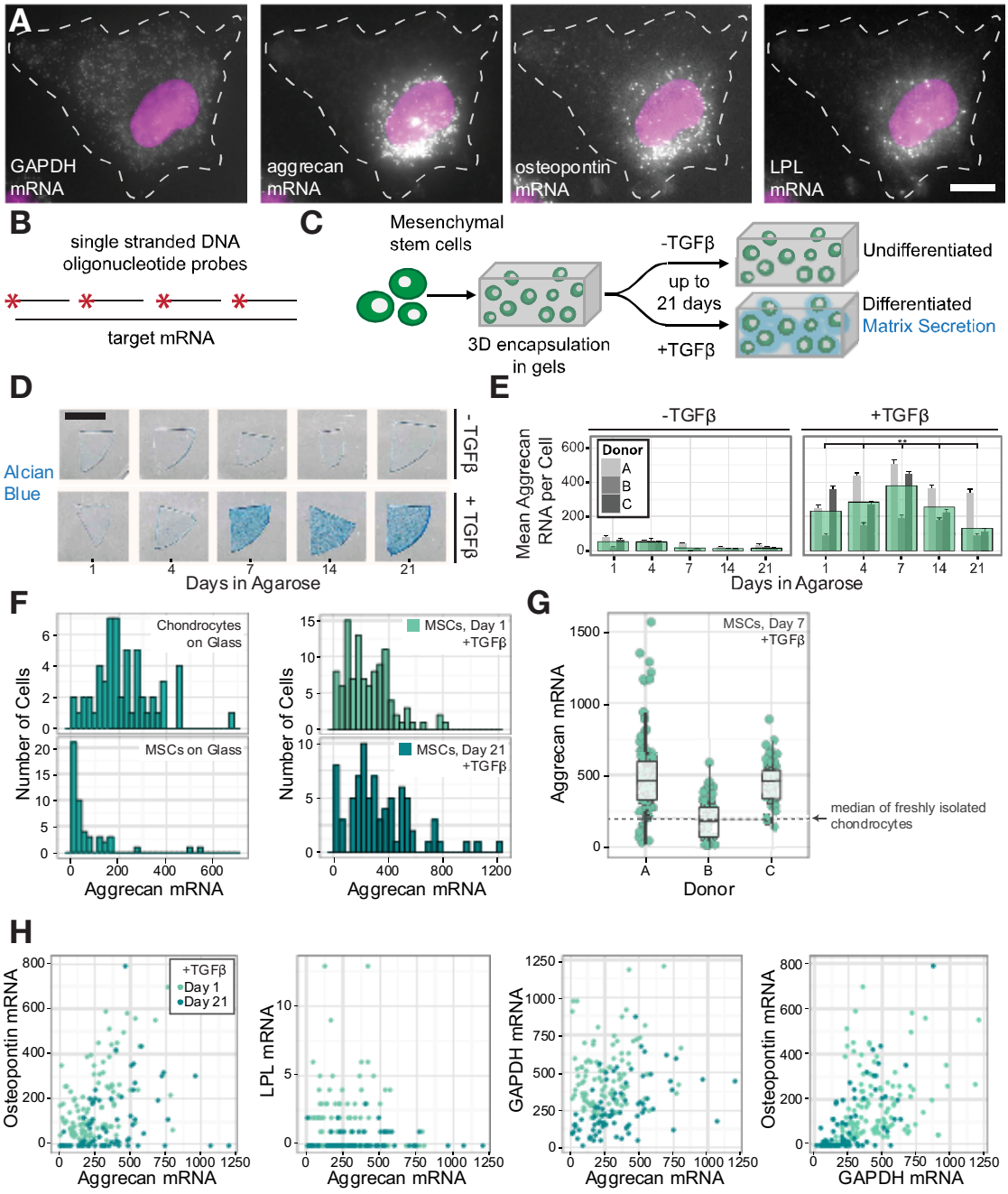

A
GAPDH mRNA
aggrecan mRNA
osteopontin mRNA
LPL mRNA
B
single stranded DNA oligonucleotide probes
target mRNA
C
Mesenchymal stem cells
3D encapsulation in gels
-TGFβ
up to 21 days
+TGFβ
Undifferentiated
Differentiated
Matrix Secretion
D
Alcian Blue
- TGFβ
+ TGFβ
Days in Agarose
E
-TGFβ
+TGFβ
Mean Aggrecan RNA per Cell
Donor
A
B
C
Days in Agarose
F
Chondrocytes on Glass
MSCs on Glass
Number of Cells
Aggrecan mRNA
MSCs, Day 1 +TGFβ
MSCs, Day 21 +TGFβ
G
MSCs, Day 7 +TGFβ
Aggrecan mRNA
median of freshly isolated chondrocytes
Donor
H
+TGFβ
Day 1
Day 21
Osteopontin mRNA
LPL mRNA
GAPDH mRNA

Figure 1 RNA FISH reveals heterogeneity in lineage marker expression in MSCSs and chondrocytes (a,b) Representative images (a) and schematic (b) of single-molecule RNA FISH, in which fluorescently labelled DNA oligonucleotides enable quantification of absolute expression of multiple genes in the same cell. Scale bar, 10 mm. (c) Chondrogenic induction scheme, involving cell encapsulation in 3D agarose constructs and exposure to TGFb. (d) Alcian blue staining for sulfated proteoglycans; Donor B shown. Scale bar, 5 mm. (e) Mean aggrecan RNA counts in MSCs cultured in 3D for up to 21 days. Narrow bars represent the mean within an individual donor; overlaid bars represent the mean across donors. Error bars indicate standard error (n = 24–128 cells per donor and condition). Means compared by t-tests with Satterthwaite approximation and simulated adjustment for multiple comparisons, **P<0.01 versus TGFb conditions, and between þ TGFb time points. See Supplementary Table 4 for all statistical comparisons. (f,g) Distributions of single-cell aggrecan expression for chondrocytes and MSCs plated on glass in basal media (f, n = 56 chondrocytes, 49 MSCs) and 3D encapsulated MSCs exposed to TGFb for 1 and 21 days (g, n = 105 cells for day 1, 79 cells for day 21; Donor A shown.) (h) Single-cell aggrecan expression for each donor after 7 days of 3D culture with TGFb relative to the median aggrecan expression in freshly isolated chondrocytes (dashed line; n = 103 cells for Donor A, 54 cells for Donor B and 65 cells for Donor C). (i) Simultaneous expression of aggrecan, osteopontin, LPL and GAPDH on day 1 and day 21; Donor A shown (n = 105 cells for day 1, 79 cells for day 21).

GAPDH copy number increased with exposure to induction media (Supplementary Fig. 1b). Thus, in aggregate, this RNA FISH analysis aligned with the canonical understanding of gene expression changes during chondrogenic differentiation (Puetzer et al., 2010). While these ensemble measures corresponded with previous findings, they did not provide information on cell-to-cell variability in expression of these lineage markers. Thus, we measured mRNA copy number on a cell-by-cell basis under baseline conditions and with differentiation. We assayed four conditions: naive MSCs in expansion culture, MSCs differentiating in engineered constructs after 1 and 21 days in induction media, and as a positive control, fully differentiated primary chondrocytes (Fig. 1f,g). For each of these groups, single-cell analysis showed striking heterogeneity in expression, with aggrecan mRNA copy number per cell spanning three orders of magnitude (10^0–10^2). Consistent with the notion that stem cells exhibit greater variability than differentiated cells, naive MSCs showed the greatest heterogeneity in aggrecan expression (as measured by the coefficient of variation, Table 1), and the coefficient of variation decreased with exposure to induction media. However, the variability remained high even after long periods of time in differentiation culture (Fig. 1g). Fully differentiated chondrocytes had the most homogeneous aggrecan expression of all the cell types and conditions we examined,

though their mean aggrecan copy number was slightly lower than that of differentiated MSCs. These data show that MSCs exhibit substantial cell-to-cell expression heterogeneity and that, while chondrogenic culture promotes a chondrocyte-like gene expression pattern, copy number remains highly variable between cells. Indeed, this variability within a population of differentiated MSCs overshadowed differences in mean expression between donors (3–4 orders of magnitude versus a maximum of twofold difference, Fig. 1h). This heterogeneity may either reflect different subpopulations that have adopted distinct fates or appear in cells that remain uncommitted. In the former scenario, if differentiated MSCs can express markers for only one fate at a time, then alternate lineage commitment should manifest as an anti-correlation between aggrecan and other lineage markers at the single-cell level. To determine whether this was the case, we performed RNA FISH for aggrecan, osteopontin and LPL in the same cells, with the latter two markers indicating osteogenic and adipogenic lineages, respectively. Rather than identifying subpopulations that were distinctly chondrogenic or osteogenic, we instead observed a slight positive correlation between aggrecan and osteopontin (Fig. 1i, Day 1 $r^2 = 0.49$, $P < 0.001$; Day 21 $r^2 = 0.34$, $P < 0.005$). Conversely, LPL expression was minimal, and did not correlate with either aggrecan or osteopontin expression (Supplementary Table 2). These data suggested that heterogeneity in marker expression after differentiation is not due to alternate lineage commitment, but rather highlights the fact that even differentiated MSCs can express high levels of markers for inappropriate lineages.

RNA levels poorly predict single-cell functional potential.

On the basis of this tremendous cell-to-cell heterogeneity in chondrogenic gene expression, we next asked whether aggrecan or other markers might serve as a means for separating robustly chondrogenic cells from the less chondrogenic ones in the initial heterogeneous population. For this to be possible, mRNA levels would need to correlate with chondrogenic capacity, indicated by the accumulation of a proteoglycan-rich extracellular matrix. To determine whether such a connection existed, we seeded MSCs in three-dimensional (3D) culture and induced chondrogenesis for 7 days, the point at which mean aggrecan expression peaked. We then performed immunofluorescent staining for aggrecan core protein (a central component of the cartilage-like extracellular matrix) simultaneously with RNA FISH using one of two probe sets: probes for markers of multiple fates (aggrecan, osteopontin, LPL and GAPDH; Batch 1 samples) or probes for chondrogenic markers (Sox9, cartilage oligomeric matrix protein (COMP) and GAPDH; Batch 2 samples). We designated cells with evidence of extracellular staining for aggrecan core protein as 'high-performing' (comprising 12–62% of the population, depending on donor), and cells lacking extracellular staining as 'low-performing' (Fig. 2a,b).

Surprisingly, aggrecan expression did not strongly predict aggrecan core protein accumulation. Indeed, even within a single donor, the distribution of aggrecan mRNA abundance in high- and low-performing cells overlapped substantially (Fig. 2c). The mRNA/cell distributions of other chondrogenic markers (COMP, Sox9), markers of alternative fates (osteopontin, LPL) and the housekeeping gene GAPDH (Supplementary Fig. 2a) also demonstrated similar overlap. While in aggregate, the high-performing cells had a greater mean expression of aggrecan, COMP and Sox9, and lower mean expression of osteopontin than low-performing cells, the magnitude of these differences was small and similar to the shift seen in GAPDH expression (Supplementary Fig. 2a,

Table 1: Mean and Coefficient of Variation Associated with Aggrecan RNA Count in Undifferentiated and Differentiated Cells		
	Mean Aggrecan	Aggrecan CV
Naive MSCs	69	1.60
Day 1 MSCs in gels	247	0.69
Day 21 MSCs in gels	334	0.72
Chondrocytes	225	0.57

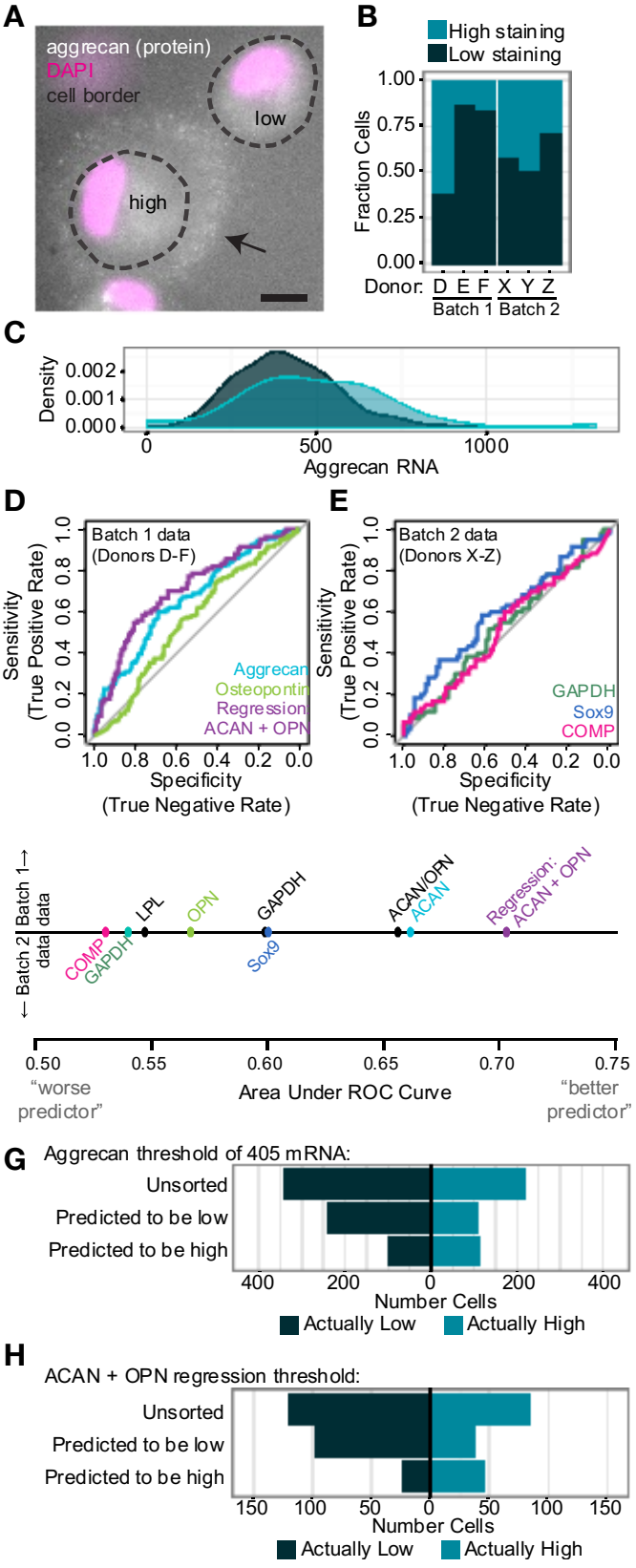


Figure 2 | Marker gene expression is a poor predictor of cartilage-like matrix production in individual MSCs. (a) Aggrecan core protein identified by immunostaining of MSCs showing high or low cartilage-like matrix formation after 7 days of 3D culture with TGFb. Scale bar ¼ 10 mm.

(b) Fraction of cells classified as high- or low-performing based on aggrecan protein staining (cells/donor: D: 78, E: 89, F: 51, X: 62, Y: 43, Z: 47).

(c) Distribution of aggrecan copy number in high- and low-performing MSC populations; probability density curve for Donor E shown (n ¼ 153 cells). (d,e) Receiver operating characteristic curves using individual gene expression and regression analysis on combinations of genes from batch 1 (d) and individual gene expression from batch 2 genes (e) to distinguish between high- and low-performing MSCs (cells/donor: D: 132, E: 153, F: 122, X: 57, Y: 42, Z: 47). (f) Summary graph of area under the curve of receiver operating characteristic curves for individual gene expression, gene expression ratios, and regression analysis of combinations of gene. (g,h) Simulated sorting of MSCs into anticipated high- and low-performing cells, using the optimized threshold of 405 aggrecan mRNA copies (g) and the optimized threshold from the ACAN þ OPN regression (h).

aggrecan: 1.35-fold increase, COMP: 1.14-fold increase, Sox9: 1.33-fold increase, GAPDH: 1.17-fold increase, osteopontin: 1.22-fold decrease). We also determined the expression ratios relative to commonly used normalization genes (that is, aggrecan/GAPDH) or to genes indicating alternate lineage specification (that is, aggrecan/osteopontin). These metrics also showed substantial overlap and small effect size (Supplementary Fig. 2a). Thus on this qualitative basis, neither absolute nor normalized single-cell expression of marker genes was highly predictive of chondrogenic capacity at the single-cell level.

To quantify the ability of transcript abundance to predict the extent of a cell's matrix accumulation, and thus sort high- from low-performing cells, we constructed receiver operating characteristic (ROC) curves to determine the 'true positive' (sensitivity) and 'true negative' (specificity) rates associated with potential mRNA thresholds. We pooled data across donors assayed using the same probes (batches 1 and 2). Within each batch, we plotted the high/low classification performance of individual genes, gene expression ratios and linear combinations of gene expression levels (Fig. 2d–f, Supplementary Fig. 2b). While each metric discriminated between high- and low-performing cells better than random chance (represented by the diagonal line on the ROC plots, and an area under the curve > 0.5), the improvements in selection specificity were relatively small. Of the individual RNA types indicative of the chondrogenic lineage, aggrecan and Sox9 were best able to discriminate between high- and low-performing cells. For example, consider the optimized threshold of 405 aggrecan mRNA, which maximizes the Youden J statistic (sensitivity þ specificity 1). Conceptually, we can designate all cells with 4405 aggrecan RNA as anticipated high performers, and others as anticipated low performers. For the donors studied, this unsorted population was 34% high- and 66% low-performing cells.

Sorting based on this optimized aggrecan threshold misclassified 37% of all cells (that is, percent of high cells predicted to be low, or low cells predicted to be high). 50% of high-performing cells were lost due to incorrect classification as ‘anticipated-low’ cells, and the fraction of high-performing cells in the ‘anticipated-high’ population was enriched only 35% over the unsorted population (Fig. 2g). A logistic regression model combining aggrecan and osteopontin expression improved on this performance only slightly, where its optimized threshold yielded a 33% misclassification rate, and enriched the fraction of high cells by 37% (Fig. 2h, versus 35% for aggrecan alone). Of the gene expression ratios, aggrecan/osteopontin was a better discriminator than aggrecan/GAPDH, though its selection performance did not surpass that of aggrecan alone. Sorting on a donor-by-donor basis was similarly ineffective (Supplementary Fig. 2b). Thus, sorting cells based on expression of aggrecan, other common differentiation markers, and linear combinations thereof would result in only marginal enrichment of the population, while substantially reducing available cell number.

Whole transcriptome profiling suggests few combinations of markers are predictive of functional potential

On the basis of the inability of aggrecan and other lineage-specific markers to robustly predict matrix accumulation at the single-cell level, we next utilized high-throughput RNA sequencing to determine whether other features of the transcriptome, and specifically factors present in the undifferentiated population, might prospectively identify MSCs with high differentiation potential. We expanded single-cell-derived MSC colonies in monolayer and collected a fraction of the cells for RNA sequencing and subsequent transcriptome

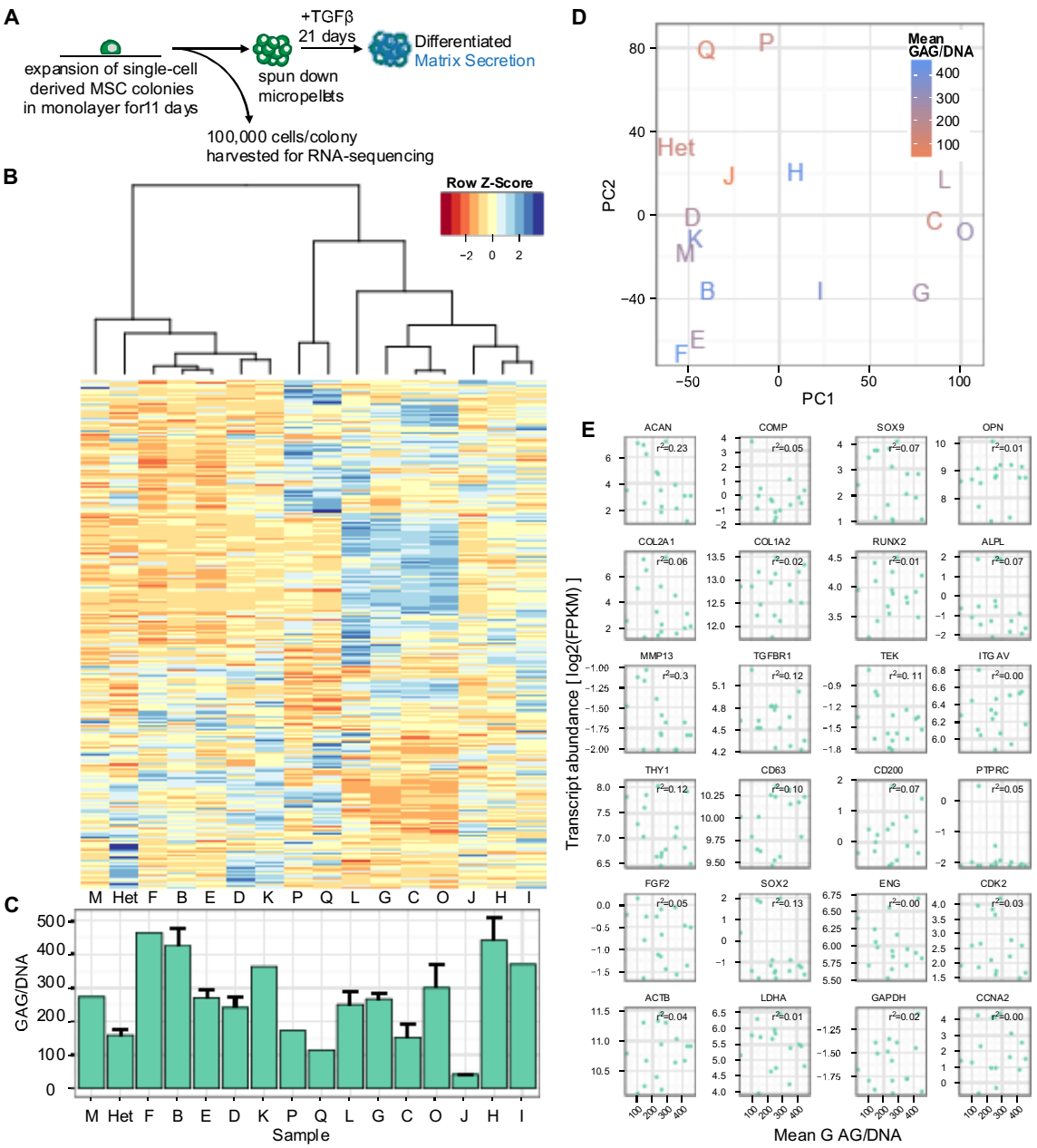


Figure 3 | Genome-wide transcriptome profiling does not predict MSC functional potential. (a) Schematic for RNA sequencing and testing of functional capacity of single-cell-derived clones. (b) Unbiased clustering of clones (and heterogeneous population) based on fragments per kilobase of transcript per million reads (FPKM) of RNA sequencing results (subsetted for genes where at least one sample had FPKM <1). (c) Glycosaminoglycan deposition per DNA in micropellets derived from clonal or heterogeneous populations (from part b) cultured for 21 days in chondrogenic (TGFb þ) culture media.For clones with limited cell number, one pellet was formed and assayed (Clones M, F, K, P, Q, W, I). For clones with cell number sufficient for multiple pellets, error bars indicate standard deviation (3 pellets—Het and clones B, E, L, G, C, O, J, H; 2 pellets—clone D). (d) Principal component analysis of same RNA sequencing results as in part b, coloured by GAG/DNA for each clone. (e) Log2 transformed FPKM of selected genes from RNA sequencing results as a function of GAG/DNA for each clone.

analysis. The remaining fraction was expanded through an additional passage, formed into pellets and cultured in the presence of TGF-ß for 21 days to assay chondrogenic potential (Fig. 3a). This evaluation of baseline MSC gene expression in clonal populations derived from single cells had the potential to identify markers that could be used to sort freshly isolated MSCs based on their gene expression signatures. An initial comparison of differential expression between clones (Fig. 3b), as compared with the deposition of extracellular matrix components of each clone (Fig. 3c), revealed no striking patterns of gene expression that correlated with subsequent matrix deposition. We also used principal component analysis to determine whether the variation between the gene expression of each clone could be used to predict functional capacity, but there was no relationship between clustering in either of the first two principal components and matrix deposition (Fig. 3d). Given that the full transcriptome lacked global predictive capacity, we next sought to broaden our conclusions from the FISH experiments by examining the sequencing data associated with individual genes. We selected a small subset of genes that corresponded to four categories of markers identified in previous studies: chondrogenic markers, stemness markers, cell cycle-associated genes and housekeeping genes (Chan et al., 2015). Consistent with our single-cell analysis results, none of these genes correlated strongly with functional potential on a clonal basis (Fig. 3e). Even the most predictive genes, MMP13 and aggrecan, correlated only weakly ($r^2 = 0.3$, $P = <0.05$ and $r^2 = 0.23$, $P = 0.062$, respectively). Altogether, this transcriptomic analysis suggests that there is no expression signature at the RNA level that could pre-identify specific clones with high chondrogenic potential.

Marker heterogeneity emerges rapidly after cell division.

On the basis of the inability of transcript levels to robustly predict matrix forming potential, we next asked whether it was propagated through cell division; that is, whether cells with a higher expression level would transmit this feature to their daughter cells. As an initial assay, we measured aggrecan copy number in every cell located within a series of small MSC colonies stimulated with TGFb (where each individual colony likely arose from a single cell, Fig. 4a). Results from this analysis showed that aggrecan copy number varied more within a single colony than it did between colonies (Fig. 4b). This result suggests that with just a few cell divisions, aggrecan levels rapidly devolved to recapitulate the heterogeneity present in the bulk population. In contrast, GAPDH was less variable than aggrecan within each colony (lower coefficient of variation, Supplementary Table 3), but showed greater differences in mean level between colonies (Fig. 4c). Thus, not every gene demonstrated the high intra-colony variability observed in aggrecan expression, and some genes were differentially expressed between colonies. However, without live cell time-lapse measurements of the cellular lineage, it was difficult to directly show that variability in aggrecan mRNA levels arose through randomization rather than heritable differences. To overcome this limitation, we next continuously tracked MSCs as they migrated and divided in induction media by live cell microscopy for 3 days, and correlated terminal aggrecan expression between sister cells with respect to the time since their last division (Fig. 4d). Shortly after division (< 12 h), sister cells had comparable aggrecan and GAPDH levels (Fig. 4e,f, Supplementary Fig. 3a,b), suggesting symmetric partitioning of RNA. However, after more than B12 h since division, sister cells showed increasingly divergent

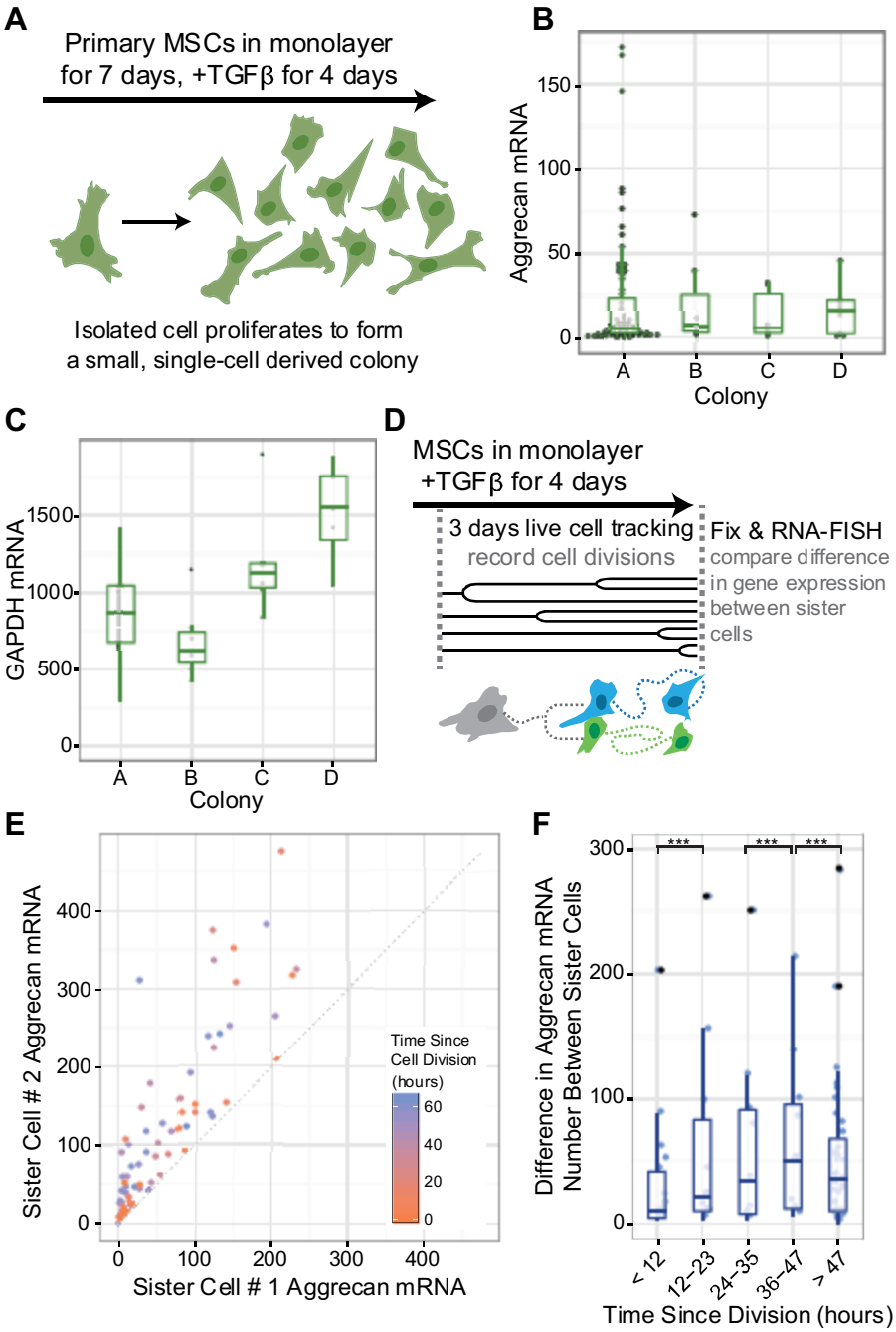


Figure 4 | Marker expression heterogeneity emerges shortly after cell division. (a–c) Gene expression in small MSC colonies. (a) Colony formation scheme. (b,c) Aggrecan and GAPDH expression in four colonies established from a single donor (n 1⁄4 75 cells in colony A, 7 cells in colony B, 6 cells in colony C, 8 cells in colony D). (d) Live cell tracking scheme to identify sister cell pairs at various times post-cell division. (e,f) Divergence in aggrecan gene expression between sister cells as a function of time since last division (n 1⁄4 81 sister cell pairs). Box hinges denote the first and third quartiles. Whiskers extend from the hinges to the most extreme data points within (1.5 * interquartile range) of the hinges. Means compared by t-tests with Satterthwaite approximation and simulated adjustment for multiple comparisons, ***P<0.001, see Supplementary Table 5 for all comparisons.

levels of aggrecan and GAPDH expression (Fig. 4e,f, Supplementary Fig. 3a,b, Supplementary Table 6). Within cell pairs, aggrecan and GAPDH divergence only weakly correlated, suggesting that the relative difference between sister cells was not globally regulated, underscoring the fact that aggrecan and GAPDH do not necessarily change together (Supplementary Fig. 3c,d). These findings may reflect a difference in cell function as a consequence of asymmetric cell division (that is, sister cells have different target expression levels) or could simply identify how asynchronous dynamic fluctuations lead to temporal differences in expression level. In either case, these differences suggest that a sorted population of high-aggrecan cells would not remain so for more than a couple of days, and may explain why, at the single-cell level, cells with high-aggrecan RNA expression are not necessarily the cells with the greatest amount of matrix deposition.

Marker genes do not identify a chondrocyte phenotype.

While aggrecan gene expression did not correlate with matrix deposition in MSCs, it is a canonical feature of the differentiated chondrocyte 'phenotype' and is widely considered to be a leading indicator of cartilage-specific extracellular matrix deposition (for example, aggrecan core protein) (Barry et al., 2001; Mackay et al., 1998; Vertel, 1995). It is also well accepted that, on serial passaging and expansion in monolayer, chondrocyte matrix production decreases along with a multi-fold decrease in the aggrecan/GAPDH ratio (Fig. 5a) (Cheng et al., 2012; Darling and Athanasiou, 2005; Elima and Vuorio, 1989; Lin et al., 2008; Ma et al., 2013). This change in expression is associated with increases in cell size and proliferation rate (Glowacki et al., 1983; Kumar and Lassar, 2009; Schiltz et al., 1973). To reconcile our finding of discordant aggrecan expression and matrix deposition in MSCs with these classical experiments that define the chondrocyte 'phenotype', we performed

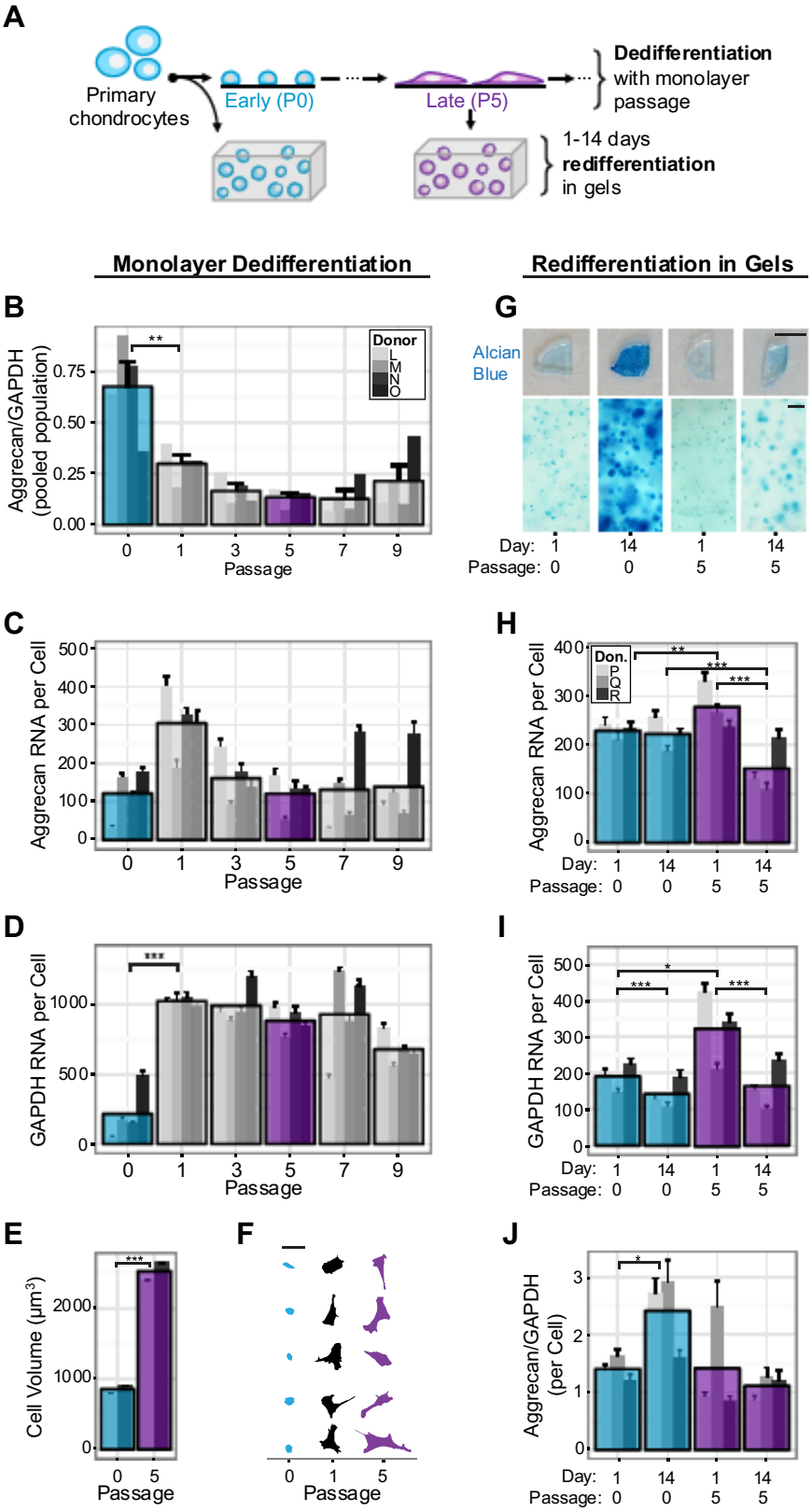

A
Primary chondrocytes
Early (P0)
Late (P5)
Dedifferentiation with monolayer passage
1-14 days redifferentiation in gels
Monolayer Dedifferentiation
Redifferentiation in Gels
B
Aggrecan/GAPDH (pooled population)
Donor
L
M
N
O
Passage
G
Alcian Blue
Day:
Passage:
C
Aggrecan RNA per Cell
H
Don.
D
O
R
D
GAPDH RNA per Cell
I
E
Cell Volume (μm³)
F
J
Aggrecan/GAPDH (per Cell)

Figure 5 | Chondrocyte de- and re-differentiation are not driven by altered absolute aggrecan expression. (a) Chondrocyte de-differentiation and re-differentiation scheme. (b–f) Analysis of chondrocytes de-differentiating with passage in monolayer culture. (b) RNA FISH counts of aggrecan pooled over the population and normalized to GAPDH expression. (c) Absolute aggrecan expression with passage number. (d) Absolute GAPDH expression with passage number (n 1⁄4 39–113 cells per donor per passage). (e,f) Chondrocyte suspended cell volume (e) and morphology (f) with passage (n 1⁄4 274–543 cells per donor per passage). Scale bar, 50 mm. (g–j) Analyses of early passage (P0) and late passage (P5) chondrocytes re-differentiating in 3D culture. (g) Alcian blue staining after 1 and 14 days of 3D culture. Top scale bar, 5 mm, bottom scale bar, 100 mm. (h) Absolute aggrecan expression. (i) Absolute GAPDH expression. (j) Single-cell aggrecan expression normalized to GAPDH expression. (n 1⁄4 46–65 cells per donor per condition). Narrow bars represent the mean within an individual donor; overlaid bars represent the mean across donors. Error bars indicate standard error. RNA count means compared by t-tests with Satterthwaite approximation and simulated adjustment for multiple comparisons. Pooled aggrecan/GAPDH expression data, cell area data, and cell volume data compared using a one-way analysis of variance with Tukey's post-hoc test. *Po0.05, **Po0.01 and ***Po0.001, see Supplementary Tables 7 and 8 for all comparisons.

RNA FISH on chondrocytes that were serially passaged in monolayer to induce 'de-differentiation' and after subsequent 're-differentiated' in 3D culture (where one would expect a resumption of the cartilage phenotype) (Benya and Shaffer, 1982; Bonaventure et al., 1994). For de-differentiation studies, we serially passaged chondrocytes nine times in monolayer with analysis at every other passage via RNA FISH. Consistent with classical findings (Benya and Shaffer, 1982; Stokes et al., 2001), the normalized ratio of aggrecan to GAPDH expression level decreased with passage number (Fig. 5b). However, and quite surprisingly, this change was not due to a decrease in absolute aggrecan copy number (Fig. 5c). Rather, aggrecan copy number showed a small but significant increase from passage 0 (initial plating) to passage 1, before returning to passage 0 mean copy number at later passages. In contrast, there was a rapid increase in mean GAPDH copy number over the first passage (increasing > fourfold) that remained at these elevated levels through additional passages (Fig. 5d). Previous studies from our group have shown that global transcription (including expression of GAPDH and many other abundant 'housekeeping' genes) correlates with and can be dictated by cell size (Padovan-Merhar

et al., 2015). We also found that chondrocyte spread-cell area generally increased with passage number (Fig. 5f, Supplementary Fig. 4) and that the mean volume of suspended cells increased by > threefold between primary isolation (passage 0) and passage 5 (Fig. 5e). Taken together, these findings suggest that aggrecan expression does not decrease with chondrocyte de-differentiation and does not correlate with chondrocyte functional potential at the population level. Instead, normalization to housekeeping genes obscures relatively minor changes in aggrecan gene expression that occurs during chondrocyte 'de-differentiation'. These single-cell data suggest that canonical markers of the chondrocyte phenotype do not accurately describe the molecular profile of de-differentiation. To further explore how normalization may confound our interpretation of gene expression changes, we forced the re-differentiation of culture-expanded chondrocytes that had lost their 'phenotype'. To do so, we encapsulated chondrocytes at early and late passage (passage 0 and 5, respectively) in 3D agarose hydrogels, and monitored matrix synthesis and gene expression over 2 weeks via Alcian blue staining and RNA FISH (Fig. 5a). Consistent with classical studies (Benya and Shaffer, 1982; Bonaventure et al., 1994), early passage chondrocytes produced matrix robustly on encapsulation, while late passage (de-differentiated) chondrocytes showed a significant attenuation in matrix deposition (Fig. 5g). RNA FISH showed that after 1 day of agarose culture, late passage chondrocytes expressed more aggrecan and more GAPDH than early passage chondrocytes (Fig. 5h,i). Over 14 days, mean aggrecan levels were maintained in early passage cells, but decreased in late passage cells. In keeping with our findings in monolayer, the aggrecan/ GAPDH ratio was strongly influenced by changes in GAPDH (Fig. 5j). These data further support the finding that absolute changes in aggrecan expression levels are not responsible for the loss of phenotype observed in serially passaged chondrocytes.

Discussion

In this work, quantitative single-cell analysis of gene expression provided evidence that the abundance of mRNA markers is only weakly linked to the chondrogenic phenotype of cartilage and progenitor cells. Specifically, we found that both MSCs and chondrocytes exhibited rampant transcriptional heterogeneity. This observation was not altogether surprising for MSCs, given that a single MSC population is comprised of a heterogeneous pool of related but distinct clonal populations. However, the transcriptional heterogeneity within individual MSC colonies suggested that this overall population heterogeneity is not entirely due to the mixing of clonal populations of varying potency, but instead likely arose from random transcriptional processes. While such heterogeneity may confound the interpretation of ensemble measurements, if this variation reflected intrinsic differences in differentiation capacity or differentiated state, then it might be harnessed towards a productive end. That is, cell sorting based on this variability could enable selection of ‘superior' sub-populations for therapeutic applications. For example, the expression of ‘stemness markers' such as SOX2 (Larsson et al., 2012), OCT4 (King et al., 2011) and NANOG (Lahm et al., 2015) can distinguish pluripotent cells from larger heterogeneous populations, and the expression of an early osteogenic marker enables enrichment of the stromal vascular fraction for osteogenic cells (Marble et al., 2014).

However, our data show that for naive MSCs, neither genome-wide transcriptional metrics nor the transcriptional abundance of MSC stemness and chondrogenic markers correlate with the ultimate functional capacity. Strikingly, the most predictive genes (aggrecan and MMP13) were negatively associated with chondrogenic capacity, potentially suggesting that high-transcriptional promiscuity in naive MSCs reflects an inability to undergo robust lineage commitment. Furthermore, our single-cell studies

showed that while naive MSCs and chondrocytes represent opposite ends of the differentiation spectrum, their absolute expression of canonical differentiation markers largely overlapped. When we monitored gene expression and cartilage-like matrix accumulation simultaneously on a cell-by-cell basis, marker expression taken at a single time point only weakly associated with cell output of extracellular matrix. Thus, we conclude that marker expression would only enable a slight enrichment of the population (~35% increase in high-performing cells over the unsorted population) while drastically restricting available cell number for therapeutic application.

One possible explanation of the disconnect between an individual cell's transcript abundance and differentiated state is that, for many genes, transcription is a stochastic process comprised of long 'silent' periods punctuated by short transcriptional bursts (Chubb et al., 2006; Golding et al., 2005; Raj et al., 2006; Raj and van Oudenaarden, 2009; Suter et al., 2011; Zenklusen et al., 2008). Bursting kinetics are strongly dependent on both the gene in question and the stimulus that is applied (Dar et al., 2012; Octavio et al., 2009; Raj et al., 2006; Suter et al., 2011), along with the position in cell cycle and cell volume (Padovan-Merhar et al., 2015; Zopf et al., 2013). For instance, stimulation (for example, TGFβ) can induce a synchronized initial burst of target gene expression, but subsequent bursts are typically asynchronous (Cai et al., 2008; Gandhi et al., 2011; Molina et al., 2013; Shah and Tyagi, 2013). Thus, two cells with fluctuating but equivalent gene expression over time may exhibit different copy number when sampled at a single time point. As recently reported (Kumar et al., 2014; Molina et al., 2013), the rate of fluctuation (slow versus fast) of a single gene manifests in the heterogeneity observed between and within small clonal clusters. Our findings of high intra-colony variability and sister cell divergence in MSCs suggest that marker copy number fluctuates rapidly over a short

timescale. As a result, absolute marker gene expression is not strongly heritable in MSCs, and we speculate that cells sorted on the basis of such expression will undergo transcriptional shifts over time and with further population expansion. In other systems, such stochastic variation in gene expression not only marks but can also determine cell fate (Choi et al., 2008; Eldar and Elowitz, 2010; Maamar et al., 2007; Raj et al., 2010; Raj and van Oudenaarden, 2008; Süel et al., 2007; Wernet et al., 2006). Here it is surprising that for aggrecan, a gene whose product plays such a critical role in the extracellular matrix, such emergent heterogeneity in transcript abundance does not appear to reflect true variation in potency.

The disconnect between expression and functional capacity (matrix accumulation) may also reflect the time history of the system and the influence of other regulatory mechanisms. Aggrecan core protein undergoes co- and post-translational modifications, and may be subject to processing or secretory errors (Luo et al., 2000; Vertel, 1995). It may be that not every cell that produces core protein can appropriately modify the core and secrete it into the extracellular space. Furthermore, integration and retention of aggrecan core protein within the extracellular matrix relies on association with the hyaluronic acid and collagen network and other molecules (Knudson, 1993; Vertel, 1995), and even aggrecan that has been integrated into the established matrix may ultimately be degraded by aggrecanases produced locally (Caterson et al., 2000). Deficiencies in any of these steps could decouple even temporally constant aggrecan mRNA expression from aggrecan core protein accumulation in the pericellular space. However, our transcriptome-wide data suggest that there is not a transcript level correlation between functional capacity and any of the genes involved in these processing steps.

Collectively, our findings in MSCs show that instantaneous aggrecan expression is only tenuously connected to matrix deposition. Moreover, differentiation of these cells fails to recapitulate the potential of native chondrocytes and does not prevent the expression of markers of alternate lineages even at the single-cell level. Our finding that chondrocyte expression of aggrecan does not decrease with de-differentiation also supports this weak connection, and raises questions as to the role of marker gene expression in defining phenotype. While aggrecan is one of the most conventional markers for the cartilage phenotype, its absolute expression did not correlate with cartilage-like matrix production and did not change as cells ‘de-differentiated'. If aggrecan expression does not change, other elements of the cell must be responsible for shifting cell fate and altering the transcriptional ‘focus' of the cell. Here our finding of major shifts in GAPDH with minor changes in aggrecan during de-differentiation suggest that de-differentiation may be better characterized as a shift in cell focus rather than a loss in specific programmatic expression of marker genes. While it is not yet clear what cell-wide changes drive this process, future work utilizing transcriptomics may identify a more comprehensive set of markers that are predictive of differentiated cell function. Until phenotype and its basis in gene expression are more precisely defined, our results suggest that it may be ineffective to design therapies that seek to bolster phenotype by increasing expression of individual genes or regulating transcriptional control of individual promoter regions, even for those genes whose products are directly related to functional matrix assembly. Simply increasing the raw RNA signal available to the cell may be insufficient, and it may also be necessary to alter the transcriptional context in which this occurs. These findings challenge the traditional notion that marker gene expression defines or is even strongly associated with

the chondrocyte phenotype and identify new directions in progenitor cell biology to establish, enforce and select subpopulations for therapeutic application.

Chapter 3: Variability in splicing from one intron to the next

Introduction

The mRNA transcribed by most eukaryotic genes are spliced, a process in which the intronic RNA sequence is removed and the exonic RNA are joined together to form the ultimate mRNA sequence. A major question in the field is how tightly associated the process of transcription and splicing are, with some work suggesting that splicing tracks closely behind the RNA polymerase as it transcribes, while other work suggests that many RNAs are fully transcribed before splicing occurs. The degree of coupling is thought to be important for the process of alternative splicing, such that different parts of the RNA are spliced together, and is a major source of transcriptomic diversity. Specifically, it is often postulated that it is necessary to completely transcribe the RNA before making a splicing "choice", implying that genes need to be spliced post-transcriptionally in order to make that choice. The fact that alternative splicing has been shown to exist for >95% of genes in mammalian cells (Pan et al., 2008) suggests that many genes may need to be spliced post-transcriptionally. The relative spatial locations of nascent pre-mRNA, fully transcribed pre-mRNA, and mature mRNA species have the potential to directly reveal where—and consequently in what order—the processes of transcription and splicing are occurring. However, to date, there have been only limited studies using molecular imaging to systematically measure the locations of these RNA intermediates. This gap has largely been due to the limits of optical resolution and the design of probes to specifically interrogate these intermediates.

In lieu of direct visualization, many have turned to biochemical fraction to infer the location of various intermediates (Bhatt et al., 2012; Drexler et al., 2019; Mayer et al., 2015; Pandya-Jones et al., 2013; Pandya-Jones and Black, 2009; Tilgner et al., 2012;

Wuarin and Schibler, 1994). Fractionation methods separate the cell into different compartments, such as the putatively chromatin associated RNA, nucleoplasmic RNA, and cytoplasmic RNA, by centrifuging the cellular components in different lysis buffers and sedimentation gradients (Mayer and Churchman, 2017). The inherent assumption made by such fractionation based methods is that the RNA species in the chromatin fraction represent nascent pre-mRNA that is still tethered to the gene body by the RNA polymerase II itself, and that once the pre-mRNA is fully transcribed, the pre-mRNA immediately moves into the nucleoplasm. Under these assumptions, any splicing observed in the chromatin fraction would be assumed to be co-transcriptional. However, this assumption may not hold: it is possible that the pre-mRNA remains in a chromatin associated compartment for some time after transcription completes, and thus splicing observed in chromatin compartment may still be post-transcriptional. Some groups have tried to test this assumption explicitly by isolating nascent RNA via metabolic labeling or the use of RNA Polymerase II antibodies, but these methods are still prone to contamination by nearby RNA. Ultimately, such alternatives are difficult to resolve without an independent and explicit verification of which RNA intermediates reside in particular compartments.

Meanwhile, advances in RNA imaging have enabled researchers to image RNA intermediates with single molecule resolution, both in fixed and living cells (Coulon et al., 2014; Levesque and Raj, 2013; Martin et al., 2013; Vargas et al., 2011; Waks et al., 2011; Zhang et al., 1994). Imaging using probes targeted both exonic and intronic regions of RNA has revealed bright nuclear foci that represent nascently transcribing RNA (Levesque and Raj, 2013; Vargas et al., 2011). The general lack of intronic signal away from these transcription sites has been taken as evidence for co-transcriptional splicing, with notable cases of post-transcriptional splicing at speckles being observed in special cases (Vargas

et al., 2011). However, these studies have a number of limitations. Firstly, owing to limitations in optical resolution, it has been difficult to visualize RNA intermediates in the peri-transcriptional region, thus making it difficult to see whether RNA are tethered to the site of transcription for splicing or dwell at the site of transcription after transcription is complete. Since splicing would appear to occur in the chromatin fraction in both of these scenarios, it is possible that much splicing is actually occurring in this peri-transcriptional region. Indeed, recent live imaging methods showed that splicing of a reporter gene happens 85% post-transcriptionally, suggesting the latter, but the use of reporter genes leaves open the question of how general such results are, given that RNA from endogenous genes may be processed differently.

Here, we designed probes to comprehensively interrogate the spatial localization of several RNA intermediates using a combination of RNA FISH and expansion microscopy (Chen et al., 2016, 2015) to reveal the ordering of transcriptional and splicing processes with single molecule resolution. We find that the portion of splicing that occurs post-transcriptionally varies from intron to intron within a single gene, but that all endogenous genes display some degree of post-transcriptional splicing. We also employ expansion microscopy to demonstrate that newly synthesized RNA dwell and undergo continuous splicing near the site of transcription after transcription is complete. These RNA are untethered to the site of transcription and eventually diffuse into either the nucleoplasm or into nuclear speckles.

Results

At least one intron of each observed endogenous gene is spliced post-transcriptionally

To directly visualize the locations of spliced and unspliced RNA relative to the transcriptional locus, we used RNA FISH to simultaneously label the exons and several individual introns of the genes of interest. Briefly, we used a series of tiled oligos conjugated to fluorescent molecules that are bound to an RNA with a complementary sequence to the oligos chosen. This allowed us to simultaneously observe introns and exons of the same gene in single cells. This revealed information about the splicing status of that gene by labeling the locations of splicing intermediates (represented by colocalized intron and exon spots) relative to the site of transcription (represented by bright colocalized intron and exon spots) and fully spliced products (represented by exon spots alone).

Visualization of splicing intermediates allowed us to distinguish between two possibilities: 1) RNA being spliced *at* the site of transcription and 2) RNA being spliced *away* from the site of transcription (Figure 6A). Observing spliced RNA at the site of transcription could represent RNA that is spliced either co-transcriptionally (tracking along with or shortly behind the polymerase) or proximally post-transcriptionally (near the site of transcription but after the process of transcription has been completed) (Figure 6A). If we observe unspliced RNA that is far away enough from the site of transcription that it could no longer be attached to the location of the gene itself, it would suggest that RNA is being spliced post-transcriptionally (distal post-transcriptional splicing, Figure 6A).

To determine what fraction of splicing occurred post-transcriptionally, we first needed to classify each spot as either a transcription site or a dispersed pre-mRNA. We computationally identified spots for both introns and exons of a particular gene, then each intron spot that was within .65uM of an exon spot was designated a colocalized exon and

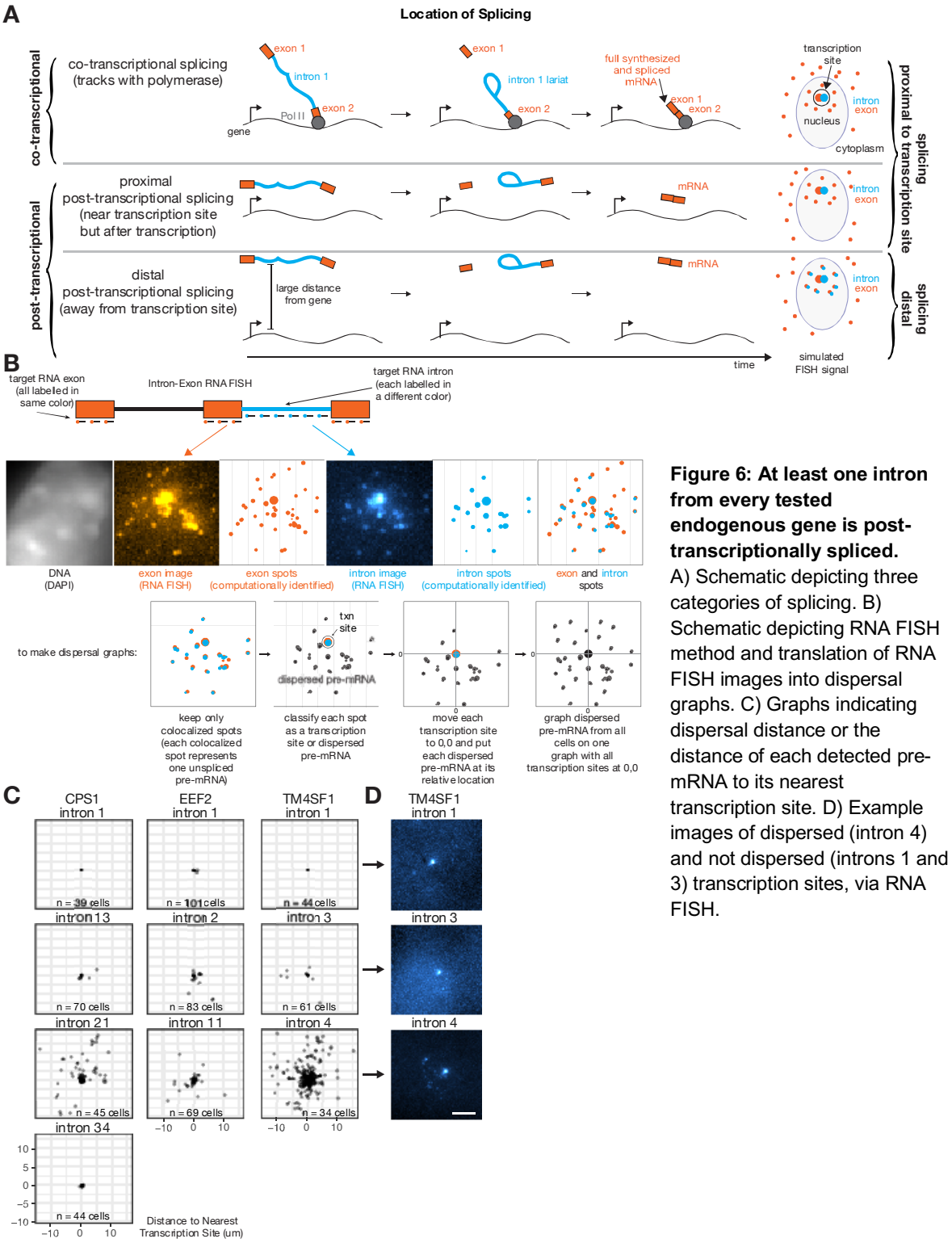


Figure 6: At least one intron from every tested endogenous gene is post-transcriptionally spliced. A) Schematic depicting three categories of splicing. B) Schematic depicting RNA FISH method and translation of RNA FISH images into dispersal graphs. C) Graphs indicating dispersal distance or the distance of each detected pre-mRNA to its nearest transcription site. D) Example images of dispersed (intron 4) and not dispersed (introns 1 and 3) transcription sites, via RNA FISH.

intron spot (Figure 6B), which we believe represents at least one pre-mRNA (potentially several pre-mRNAs when found at the site of transcription). Intron spots that did not colocalize with an exon spot were presumed to be degradation products or non-specific background and were discarded, though these were generally <25% of intron spots (Table 2). We used an intensity threshold to categorize each colocalized exon and intron spot as either a transcription site or a dispersed pre-mRNA (Figure 6B, see Supplementary Figure 5A,B for more information about the classification of transcription sites). We then displayed the distance of each dispersed pre-mRNA from its nearest transcription site for all cells on a single graph to display all dispersal per intron (Figure 6C). We found that for the four genes tested, at least one intron displayed some splicing intermediates away from the site of transcription (3-5 introns tested per gene, Figure 6C). This suggests that some introns (for which we see splicing intermediates far from the site of transcription) are being spliced post-transcriptionally.

We made the assumption that intron containing pre-mRNA away from the site of transcription are post-transcriptionally spliced, and not degraded. This assumption holds because the number of remaining unspliced introns decreases with distance from the transcription site, suggesting that introns are excised progressively as transcripts get further away from the transcription site (Supplementary Figure 5E).

Given that individual introns within the same gene displayed differing levels of dispersal, we asked whether the degree of dispersal could depend on the order in which transcription occurs or whether the introns are being spliced independently of transcriptional order. To test this, we We didn't observe an increase in dispersal on a 5' to 3' basis (Figure 6C). This lack of increased dispersal on a per gene basis suggests that the introns are being spliced not in a 5' to 3' order (Figure 6C).

The differential dispersal of each intron within a gene suggests that each intron has a particular splicing rate which governs the time it takes until that intron is spliced, apparently independently of neighboring introns.

Transcripts are untethered to the site of transcription and dwell at the site of transcription after transcription is completed

While conventional RNA FISH allowed us to determine what portion of splicing is happening distally post-transcriptionally (Table 2, Supplementary Figure 5C), the resolution limits of conventional light microscopy made it difficult to distinguish whether splicing at the site of transcription was happening co-transcriptionally or proximally post-transcriptionally. It was impossible to distinguish whether transcripts are being spliced during the process of transcription or after the process of transcription has been completed because conventional light microscopy cannot distinguish which RNA at the site of transcription are being actively transcribed or have completed transcription. It is difficult to distinguish those two possibilities because, with conventional light microscopy, all of the RNA at the site of transcription are only visible together as one large transcription site spot.

In order to visualize individual RNAs and proteins at site of transcription, we used a combination of expansion microscopy and RNA FISH to expand the transcription site and visually identify single spots at the site of transcription, instead of a bright blob (Figure 7A, B). We knew that spots at these sites represent completed or almost completed transcripts because we observed multiple spots in a cloud from our 3' probe set, which targets specifically the 3' end of the RNA, so whenever this probe set is visible it means the end

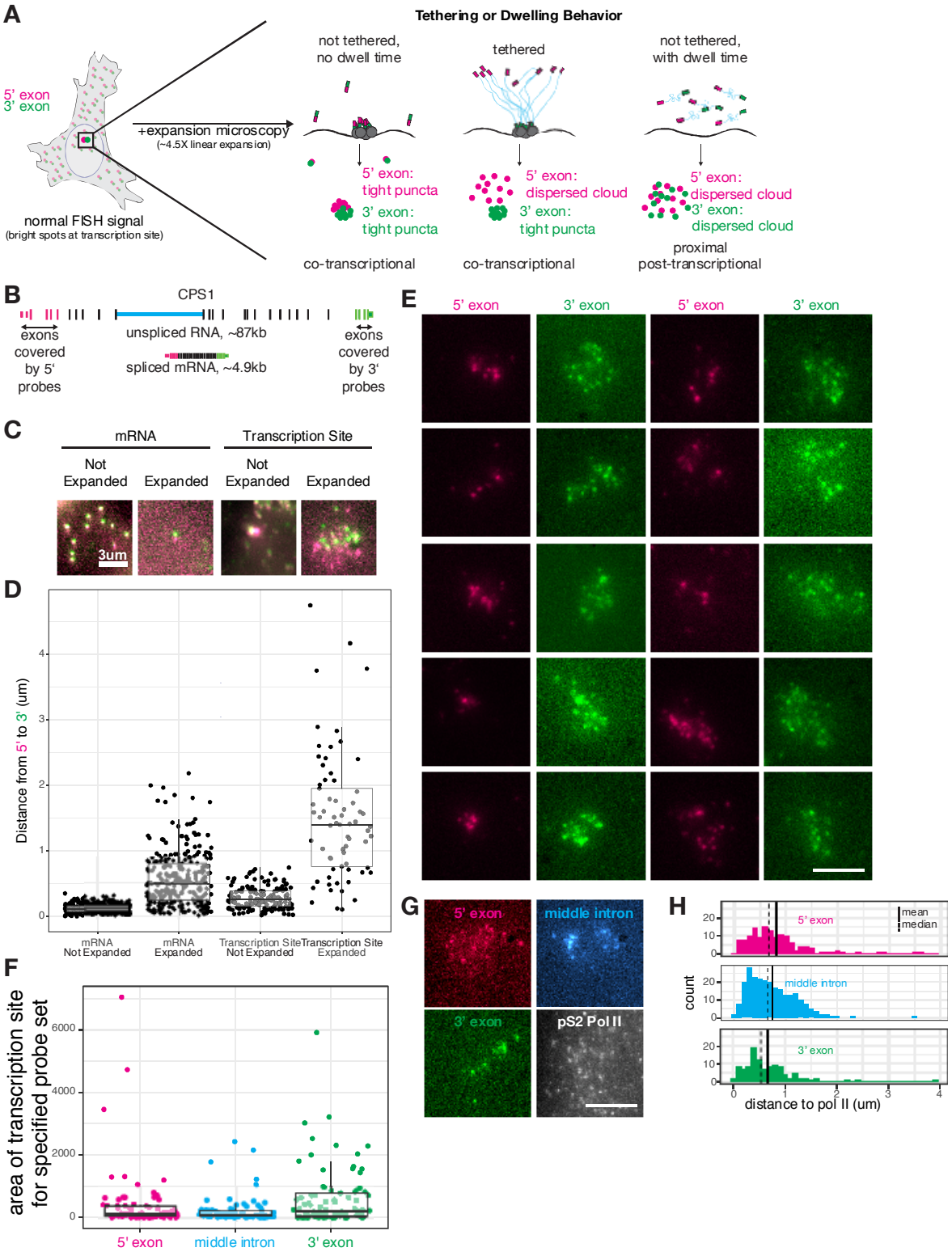

A
Tethering or Dwelling Behavior
5' exon
3' exon
+expansion microscopy
(~4.5X linear expansion)
normal FISH signal
(bright spots at transcription site)
not tethered,
no dwell time
5' exon:
tight puncta
3' exon:
tight puncta
co-transcriptional
tethered
5' exon:
dispersed cloud
3' exon:
tight puncta
co-transcriptional
not tethered,
with dwell time
5' exon:
dispersed cloud
3' exon:
dispersed cloud
proximal
post-transcriptional
B
CPS1
exons
covered
by 5'
probes
unspliced RNA, ~87kb
spliced mRNA, ~4.9kb
exons
covered
by 3'
probes
C
mRNA
Transcription Site
Not
Expanded
Expanded
Not
Expanded
Expanded
3um
D
Distance from 5' to 3' (um)
mRNA
Not Expanded
mRNA
Expanded
Transcription Site
Not Expanded
Transcription Site
Expanded
E
5' exon
3' exon
5' exon
3' exon
F
area of transcription site
for specified probe set
5' exon
middle intron
3' exon
G
5' exon
middle intron
3' exon
pS2 Pol II
H
5' exon
mean
median
middle intron
3' exon
count
distance to pol II (um)

Figure 7: Transcripts dwell at the site of transcription after transcription is completed.
A) Schematic depicting potential scenarios for RNA immediately after the process of transcription completes. B) Gene and probe position diagram for CPS1 pre-mRNA and mRNA. C) example images of RNA FISH before and after expansion microscopy for individual mRNA and transcription sites. D) Distance from 5' to 3' of expanded and unexpanded mRNA and transcription sites, as detected by CPS1 RNA FISH. E) Example images of 5' and 3' CPS1 RNA FISH after expansion. F) Quantification of E and other images, CPS1 RNA FISH after expansion microscopy. G) Example images of co-IF (polymerase II) and RNA FISH for CPS1 after expansion. H) Quantification of images like that in G, and others, representing distance from each RNA spot to the nearest polymerase II IF spot.

of the RNA is largely transcribed (Figure 7C,D). We therefore hypothesized that these are pre-mRNA or mRNA moving away from the site of transcription, but their localization to adiscrete region suggest that there is a proximal region to the site of transcription through which the RNA are moving slowly, potentially where splicing is still occurring.

In addition to these 3' clouds, we observed that the 5' and 3' ends of the pre-mRNA at the site of transcription are farther apart than their cytoplasmic counterparts (Figure 7D), suggesting that either the RNA is stretched more than cytoplasmic RNA, or they are longer because there are more introns incorporated into the RNA. The 3' spot clouds suggest two things: 1) that transcripts move slowly through a region proximal to the transcription site and are untethered to the gene during this movement, or we would expect to see tight 3' spot colocalization (Figure 7A,C) and 2) that transcripts dwell after completion at the site of transcription, or we would expect to see both tighter 5' and 3' spots (Figure 7A,C). If transcripts did not dwell at the site of transcription, we would expect to see only actively transcribed nascent RNA at the site of transcription (forming a tight cloud around polymerase II puncta), but the presence of many 3' spots suggest that there are many RNA that are not actively being transcribed.

To confirm that the RNA at the site of transcription are not actively being transcribed, we co-stained with actively elongating polymerase II (phosphoS2)(Figure 7G,H). The lack of

active transcription is confirmed by the observation that there are many 3' spots that do NOT localize with actively elongating polymerase II (Figure 7H). We observe that the mean and median distances of 3' spots to polymerase II is greater than the majority of expanded mRNA, suggesting that most spots are too far away from polymerase II to be actively being transcribed.

Taken together, this data suggests that many transcripts proximal to the site of transcription are not actively being transcribed but are instead moving slowly through a small zone around the site of transcription, where they are untethered to the site of transcription but still in the process of being spliced.

Sequencing corroborates RNA FISH dispersal results

To corroborate our RNA FISH results, we turned to sequencing of nascent RNA to determine the splicing status of RNA that has recently been transcribed and compared that to our dispersal metrics. Nascent RNA was isolated by a combination of metabolic labeling of newly synthesized RNA and isolation of RNA that cosediments with the chromatin fractions of cells. This isolated RNA was then subject to standard high throughput sequencing methods and we calculated a "splicing index" per intron that is defined as the number of spliced reads for a given junction divided by the number of unspliced reads for the same junction (Figure 8A). This splicing index represents how "co-transcriptionally spliced" a particular intron is, introns with a high splicing index are very co-transcriptionally spliced and introns with a low splicing index are very post-transcriptionally spliced.

Sequencing of this metabolically labelled and chromatin associated RNA also suggested that introns have variable splicing even within endogenous genes (Supplementary Figure 6A). Applying our splicing index to the introns that we have dispersal data for shows a variable splicing index from intron to intron (Supplementary Figure 6A). Directly comparing splicing index and mean dispersal shows an inverse relationship, as expected, introns with a high splicing index (very “co-transcriptionally” spliced) have low dispersal, and introns

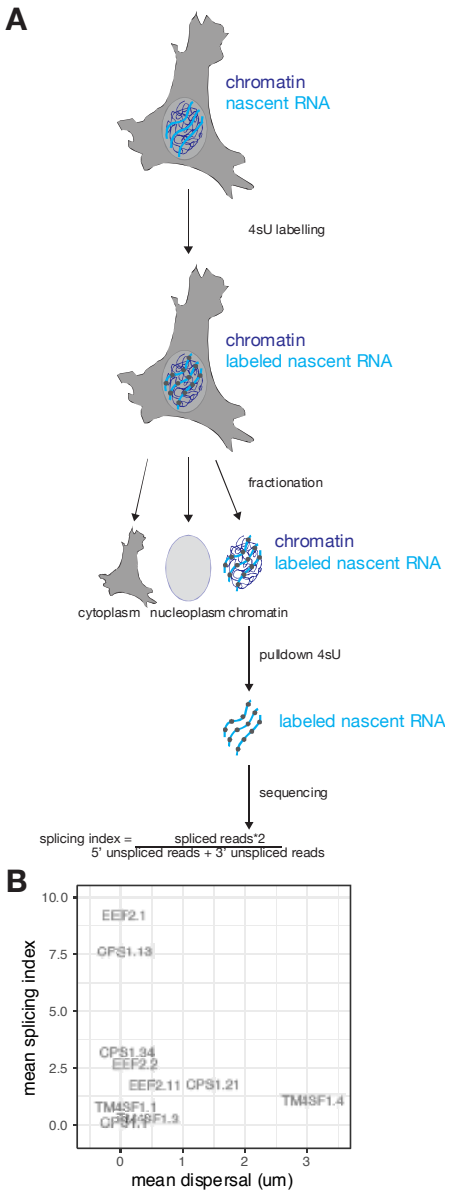


Figure 8: Sequencing corroborates RNA FISH dispersal results. A) Schematic of sequencing experiment design. B) Comparison of mean dispersal (as assessed by RNA FISH, see Figure 6) and mean splicing index (as assessed by sequencing).

with a low splicing index (very “post-transcriptionally” spliced) have high dispersal (Figure 8B). However, there are a number of introns for which both the splicing index and mean dispersal were low, suggesting that these low-splicing-index and low-dispersal introns are subject to technical noise or there may be a great deal of “proximal post-transcriptional” splicing, which would result in low dispersal but also a low splicing index.

We also analyzed transcription site size for each intron for which we have FISH data (Supplementary Figure 6B), which we anticipated would be another metric by which we can measure proximal post-transcriptional splicing because large transcription sites may represent high numbers of “dwelling” RNA that are done with transcription but being actively spliced proximal to the site of active transcription. Similar to mean dispersal, RNA FISH-based analysis of transcription site size also displays an inverse relationship with splicing index; however, there remain several ambiguous introns in the lower left quadrant of the graph that have both low splicing index (very “post-transcriptionally” spliced) and low transcription site size (Supplementary Figure 6B).

The anticorrelation of splicing index with both mean dispersal and transcription site size suggests that sequencing of nascent RNA corroborates mean dispersal and transcription site size as methods to assess the amount of post-transcriptional splicing occurring.

A fraction of unspliced pre-mRNA localize to a speckle-proximal compartment after exiting a slow moving zone proximal to the transcription site

We wondered where transcripts went after they were released from the slow-moving transcription-site-proximal zone that was revealed by expansion microscopy. We hypothesized that the transcripts could do one of three things: 1) freely diffuse away from

the transcription site, 2) be tethered to the site of transcription, 3) fill a compartment, potentially around or adjacent to the site of transcription or other nuclear bodies (Figure 9A).

However, there were relatively few unspliced RNA that made it out of that transcription-proximal zone, making it difficult to characterize the behavior of these pre-mRNA. Thus, we inhibited splicing to generate more pre-mRNAs making it easier to quantify their behavior.

Upon splicing inhibition, we observed three distinct phenotypes for pre-mRNA species, one in which there are increased dispersed pre-mRNA throughout the nucleus (nuclear dispersal genes), one in which the pre-mRNA move into a small compartment, likely to be around the site of transcription or another nuclear body (compartmentalization genes), and one in which the transcription sites look identical to transcription sites before splicing inhibition (non-responsive genes) (Figure 9B,C). The compartmentalization phenotype only appeared for three tested genes (out of 16): EEF2, GAPDH, and RPL13A (Figure 9C), whereas we observe a nuclear dispersal phenotype for 7 genes and no response for 6 genes (Figure 9H).

One of our hypotheses was that these pre-mRNA might be tethered to some location within the nucleus, whether that be the DNA that they are transcribed from or a point within the compartments that we observe. However, it is likely that these pre-mRNA are filling up a compartment and are not tethered within this compartment because we observe full colocalization of 5’ and 3’ exons within the compartment (Figure 9D), whereas we might otherwise expect a tighter colocalization of one section (5’, middle, or 3’) of the RNA if the RNA were tethered to one specific point within the compartment (Figure 9A, Supplementary Figure 7A).

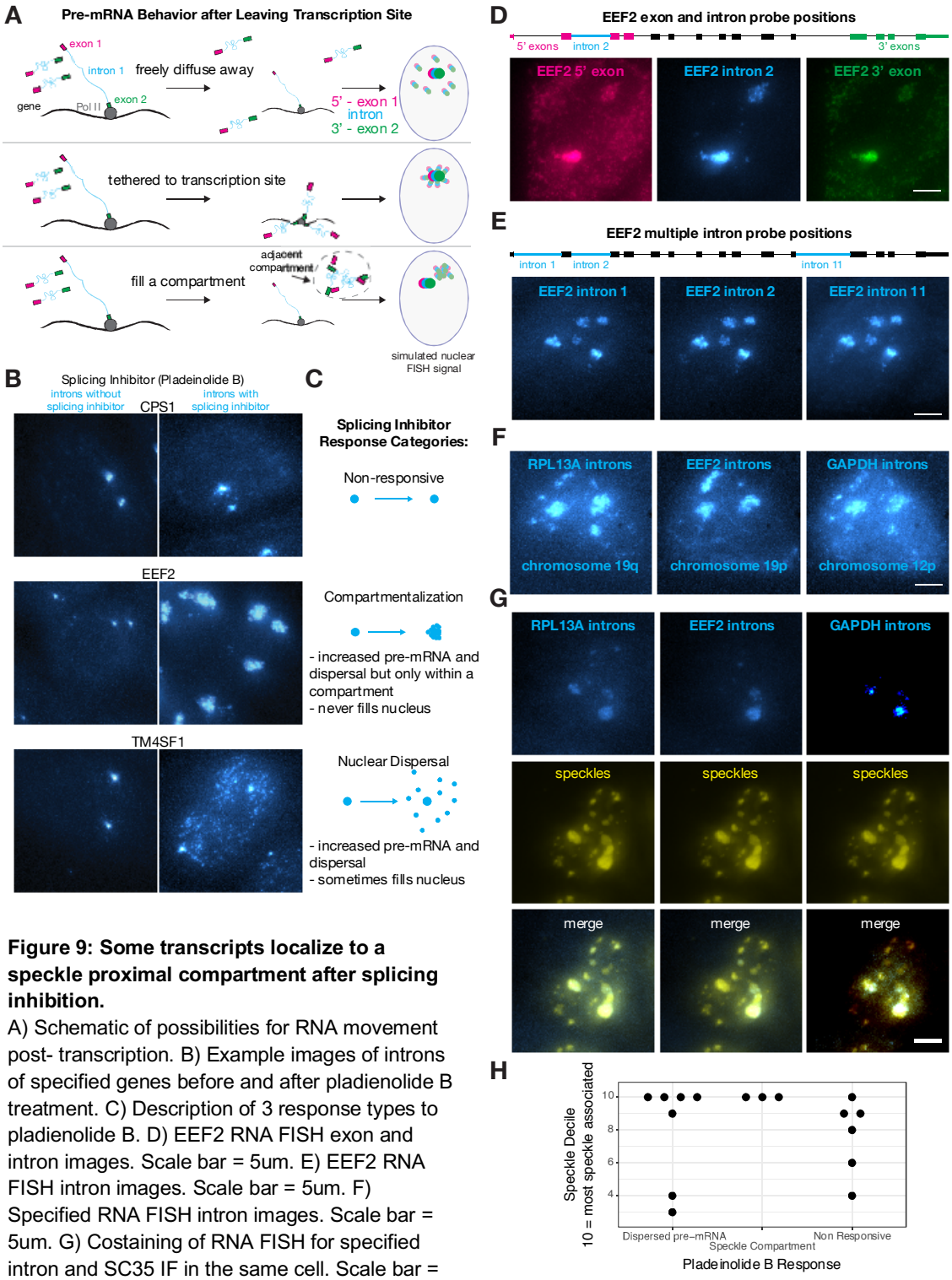


Figure 9: Some transcripts localize to a speckle proximal compartment after splicing inhibition.

A) Schematic of possibilities for RNA movement post- transcription. B) Example images of introns of specified genes before and after pladienolide B treatment. C) Description of 3 response types to pladienolide B. D) EEF2 RNA FISH exon and intron images. Scale bar = 5um. E) EEF2 RNA FISH intron images. Scale bar = 5um. F) Specified RNA FISH intron images. Scale bar = 5um. G) Costaining of RNA FISH for specified intron and SC35 IF in the same cell. Scale bar = 5um. H) Quantification of speckle decile from previously published data, compared with PlaB response.

We also observe that multiple introns from the same gene colocalize to the same compartments (Figure 9E), suggesting that all unspliced pre-mRNA from a gene localize to the same compartments. Additionally, we also observe multiple genes localize to a similar set of compartments (Figure 9F). Two of these genes are located to opposite ends of the same chromosome (EEF2 and RPL13A, chromosome 19) and the third gene is located on a separate chromosome (GAPDH, chromosome 12) (see Supplementary Figure 7B for karyotype). However, when observing the three genes before splicing inhibition, they do not appear to colocalize with one another, suggesting perhaps that the compartments they colocalize to after splicing inhibition may be somehow targeted by these genes after splicing inhibition.

We wondered if these compartmentalized pre-mRNA were located near any other nuclear structures. One candidate was nuclear speckles, which are compartments in the nucleus that contain concentrated splicing and transcription factors (Spector and Lamond, 2011). To determine whether these compartments colocalize with nuclear speckles, we performed RNA FISH simultaneously with immunofluorescence for SC35, a component of speckles, and see that these compartmentalized pre-mRNA do indeed colocalize with nuclear speckles both before after splicing inhibition (Figure 9G, Supplementary Figure 7D). This association between speckles and compartmentalized pre-mRNA is confirmed by previously published high throughput sequencing data that quantified the distance of transcripts from speckles and other nuclear compartments (Chen et al., 2018). From that data, we find that compartmentalized genes are all within the most speckle associated decile of the data, and all other tested genes (both nuclear dispersal and non-responsive) exhibit a range of distances to speckles (Figure 9H).

This data suggests that pre-mRNA from specific genes are sequestered into speckle-proximal compartments before splicing occurs.

Intron splicing rate is not an inherent trait of individual introns and can vary with transcription level

We wondered whether splicing rate (as represented here by dispersal), was an inherent property of each intron tested or whether the splicing could vary with context, such as transcription level. To test whether splicing rate was an inherent property of introns, we treated A549 cells with dexamethasone to induce transcription of the gene FKBP5, and then performed RNA FISH along varying time points in dexamethasone and assessed dispersion (Figure 10A, B). Over time, we see an increase in both exon and intron spot counts (Figure 10B), and a corresponding increase in intron dispersal (Figure 10C) but only for certain introns. The fact that dispersal increases with transcription level for these introns suggests that dispersal is therefore not an inherent property of each intron but depends on local context (Figure 10C). However, some introns do not exhibit an increase in dispersal even with long exposure to dexamethasone (8 hours) (Figure 10C), suggesting that some introns may have a splicing rate that is fast enough to still exhibit transcription site localized splicing even in case of increased transcription. This data suggests that dispersal, or the percentage of splicing that happens distally post-transcriptionally, can vary with gene context, such as transcription level, and is therefore not an inherent property of each intron.

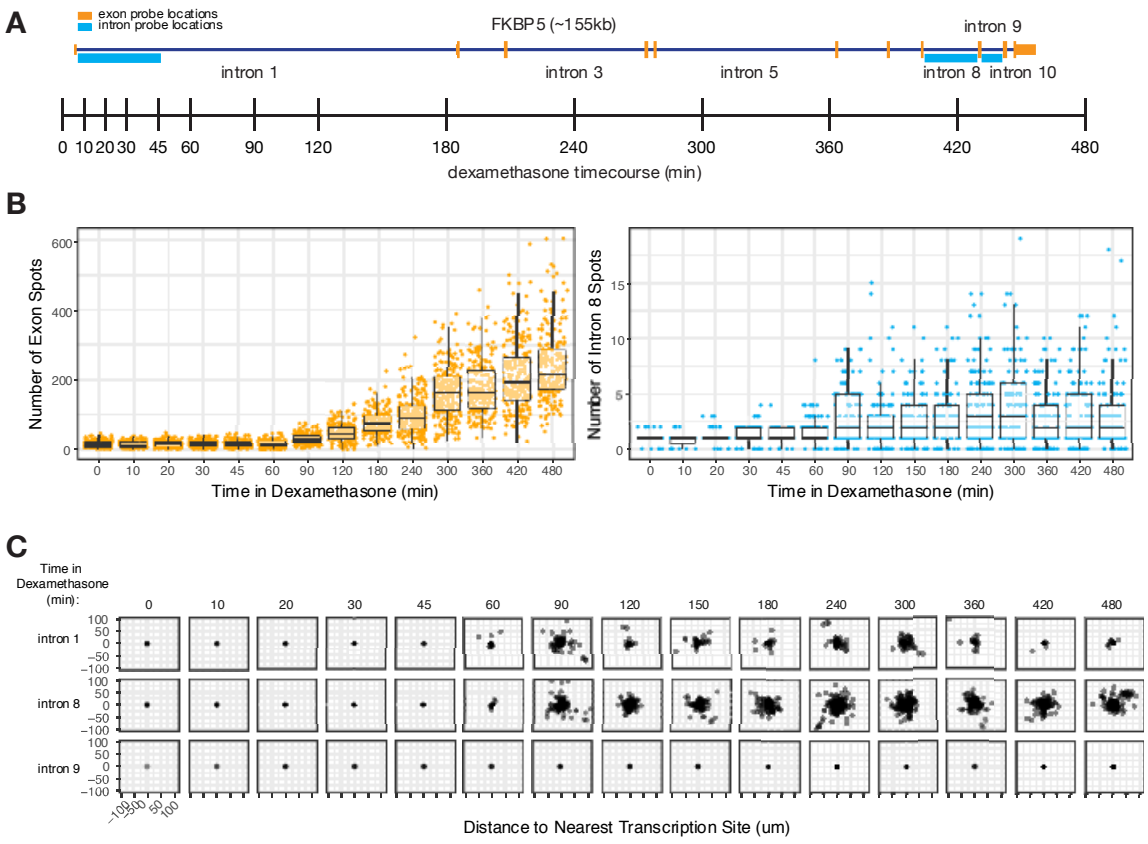


Figure 10: Intron dispersal varies with transcription level and is therefore not an inherent property of each intron. A) Gene and probe diagram for FKBP5 and schematic of dexamethasone treatment schedule. B) Quantification of FKBP5 RNA FISH exon and intron spots over time of treatment in dexamethasone. C) Dispersal graphs (as quantified from RNA FISH) of FKBP5 introns 1, 8, and 9 over time in dexamethasone.

Table 2: Exon intron colocalization and post-transcriptionality of splicing for all tested genes					
Gene	**Intron #**	**# of Probes**	**% coloc with exon**	**mean dispersal (um)**	**% of intensity that is not local to txn site**
CPS1	1/37	10	75.5814	0	0
CPS1	13/37	12	88.2716	0.067485645	1.337543
CPS1	21/37	24	74.71264	2.0966995	27.7981585
CPS1	34/37	16	88.88889	0.03229656	1.952689
EEF2	1/14	21	91.66667	0.93700431	15.69285
EEF2	2/14	21	91.73554	0.41633783	6.482196
EEF2	11/14	13	94.08867	1.67628235	24.386486
TM4SF1	1/4	17	98.59155	0	0
TM4SF1	3/4	24	94.78261	0.34731502	2.422979
TM4SF1	4/4	24	87.84648	3.00480873	47.094267
FKBP5**	1/10	32	77.52809	1.2286272	22.451107
FKBP5**	3/10	32	94.64286	0.4514456	3.692255
FKBP5**	5/10	32	73.4375	2.852407	6.836791
FKBP5**	8/10	32	81.41593	1.4870159	33.077603
FKBP5**	9/10	14	93.89671	0.5677152	3.679758
FKBP5**	10/10	11	71.60494	0	0

Discussion

Our results suggest the following model of splicing: first, the mRNA is transcribed. Splicing occurs continuously and independently between different introns while transcription is occurring and also while the pre mRNA moves slowly through a region proximal to the transcription site, after which it moves either to the speckle compartment or freely in the nucleoplasm. This is in contrast to prevailing ideas of splicing, which suggest that splicing happens immediately (within 15-20 seconds or 45-100 nucleotides of transcription) after transcription of each intron is completed (Alpert et al., 2016; Carrillo Oesterreich et al., 2016; Eser et al., 2016; Huranová et al., 2010; Martin et al., 2013; Wallace and Beggs, 2017). This contrast may be due to a variety of different reasons, including but not limited to species specific differences and the use of different assays to measure the timing of splicing (Alpert et al., 2016).

We see that introns of most endogenously expressed genes are spliced in a variety of spatial distributions, with all the genes we probed demonstrating at least one intron with some splicing occurring distally from the transcription site. This suggests that each intron has a distinct splicing rate, with at least one intron for each gene spliced partly distally post-transcriptionally. Our results also suggest that introns are spliced independently of one another, whether they are proximal or distal to each other. This is in contrast to other work which suggests splicing of particular introns is controlled or gated by the splicing of other introns or exons within the same gene (Kim et al., 2017).

One of the original models for splicing is the “first come, first serve” model in which each intron is immediately spliced upon the completion of transcription, in a 5’ to 3’ order (Aebi and Weissman, 1987). Our results suggest that this is not the case, based on seeing low splicing rates (or high dispersion) for even the 5’ most introns of some genes. This lack of

first come, first serve splicing is confirmed by others in several different situations (de la Mata et al., 2010; Kessler et al., 1993; Yang et al., 2012).

Our expansion microscopy results also suggest the existence of a region proximal to the gene through which the pre-mRNA moves slowly before being allowed to freely move through the nucleus for export. The slow movement through this region may be caused by already existing chromatin structures or may be due to a concentration of protein factors that is created by the process of transcription itself, and the transcription associated processing steps such as capping and polyadenylation. Work by Pandaya-Jones et al. and Bhatt et al. suggests that fully transcribed but unspliced RNA remain in the chromatin compartment (the RNA that cosediments with chromatin after centrifugation) and are fully spliced before release into the nucleoplasm (Pandya-Jones et al. 2013; Bhatt et al. 2012), which would be consistent with RNA moving slowly through a transcription site proximal compartment while the process of splicing is completed.

Our results also suggest that pre-mRNA are not tethered to the site of transcription while they move through this transcription site proximal compartment. This is in contrast to what is suggested by Dye et al., where they suggest that exons are tethered to polymerase II as splicing is occuring (Dye et al. 2006). The lack of colocalization of introns with polymerase II also suggests that splicing is not happening with polymerase II immediately downstream of those introns. This contrasts the data of Alexander et al., which suggest that almost all splicing occurs while polymerase II is still paused proximal to the intron that was recently transcribed (Alexander et al. 2010). This may be due to species specific differences, however, there have been many links between polymerase II pausing and splicing (for a review, see (Carrillo Oesterreich et al. 2011)).

In addition to a slow moving transcription proximal compartment and lack of an RNA-gene tether, our expansion microscopy suggests that the distance between the 5' end and the 3' end of RNA at the site of transcription is greater than the distance between the 5' end and 3' end of mRNA. This increased 5'-3' distance suggests that the transcripts at the site of transcription are either unpackaged, perhaps due to decreased RBP occupancy, or are simply longer because there are likely more introns incorporated close to the site of transcription than away from it. If this 5' to 3' distance increase is due to increased incorporation of introns, this would challenge the idea that splicing is happening immediately post-transcription.

We also wondered what happens to unspliced pre-mRNA that escape this slow-moving transcription proximal compartment. Our splicing inhibition results suggest that some genes localize to a speckle-proximal compartment after escape, and some move freely through the nucleoplasm, with some dependence on pre-splicing inhibition speckle association. The consistency of our observation with another speckle associated metric across cell types and assays suggests that the speckle association of these genes might be constitutive (Chen et al. 2018). We also observe that pre-mRNA of compartmentalization genes seem exclusively localized to this speckle-proximal compartment, perhaps as a way to sequester the unspliced RNA. It is possible that escape of these unspliced compartmentalized pre-mRNA away from the site of transcription and into the cytoplasm for translation may be particularly deleterious for a cell because those genes displaying a compartmentalization phenotype are all important genes (Curtis et al., 2010; Kapp and Lorsch, 2004; Nicholls et al., 2012). It is also possible that these RNA, because of their abundance, sequence factors, or associated RNA binding proteins, are more likely to integrate into compartments, such as nuclear speckles (Chen et al., 2018).

We also see that increased transcription level correlates with intron dispersal, which suggests that the percentage of splicing that happens distally post-transcriptionally is regulated by transcription level for at least some introns. This may also explain why we observe the post-transcriptional splicing of all genes measured as all of them were highly expressed. Work by Ding and Elowitz suggests that, for high expressing genes, splicing can act as an “economy of scale” filter, in which the expression of already highly expressed genes is amplified by increased splicing efficiency, potentially due to association with nuclear speckles (Ding and Elowitz 2019).

Taken together, our data challenge the longstanding idea that splicing is happening immediately co-transcriptionally for endogenous genes.

Conclusion

In this thesis, I address two separate questions:

1) How do we pick “good” mesenchymal stem cells?

2) Is splicing happening co-transcriptionally?

I use a variety of methods to address these questions, but most of my methods used focused on characterizing the cell-to-cell variability in transcription and splicing, respectively.

My results suggest that, despite the marked heterogeneity in marker expression in naive mesenchymal stem cells, traditional markers of chondrocyte function will not identify high performing mesenchymal stem cells initially after isolation, or even after the process of differentiation. This may be due to many factors which require further investigation, such as modulation of genes in the secretory pathway or defects in post-translational modifications of extracellular matrix proteins. Regarding variability in transcription, it is likely that much of this variation is due to the bursting nature of transcription (Raj et al., 2010, 2006; Raj and van Oudenaarden, 2008). It has been established that transcription occurs in discrete bursts, as opposed to continuously, for almost all genes (Suter et al., 2011). These bursts, when infrequent, could result in the massive variability that we see in expression levels of marker genes in mesenchymal stem cells. With infrequent bursts, cells with high expression at one moment in time may not experience another burst until all of their mRNA are degraded, leading to the observed lack of “cellular memory” that we see with recently divided mesenchymal stem cells. This lack of cellular memory may explain one reason why traditional markers do not represent high performing mesenchymal stem cells.

In addition to the variability in expression in mesenchymal stem cells, our study of variability in chondrocyte marker expression suggests that absolute expression of the markers measured also do not reflect high performing chondrocytes, especially after time in monolayer culture. This suggests that cellular context is particularly important in this case.

In the future, I would continue this work by using the transcriptome data generated with new sequencing analysis tools that might allow us to make use of the continuous variable data like GAG deposition of each clone. I would use this transcriptome data to determine which cell surface markers could be of use to sort mesenchymal stem cells out for their ultimate cartilage tissue engineering performance. I could then test this by actually sorting and differentiating mesenchymal stem cells. In addition, I believe it is necessary to study several other aspects of cartilage extracellular matrix deposition to determine where the breakdown occurs between expression of traditional markers and their ultimate deposition. Several likely candidates include post-translational modifications of these proteins, as aggrecan especially is heavily glycosylated post translationally, and defects in the secretory pathway, which may prevent proteins produced from being deposited into the extracellular space.

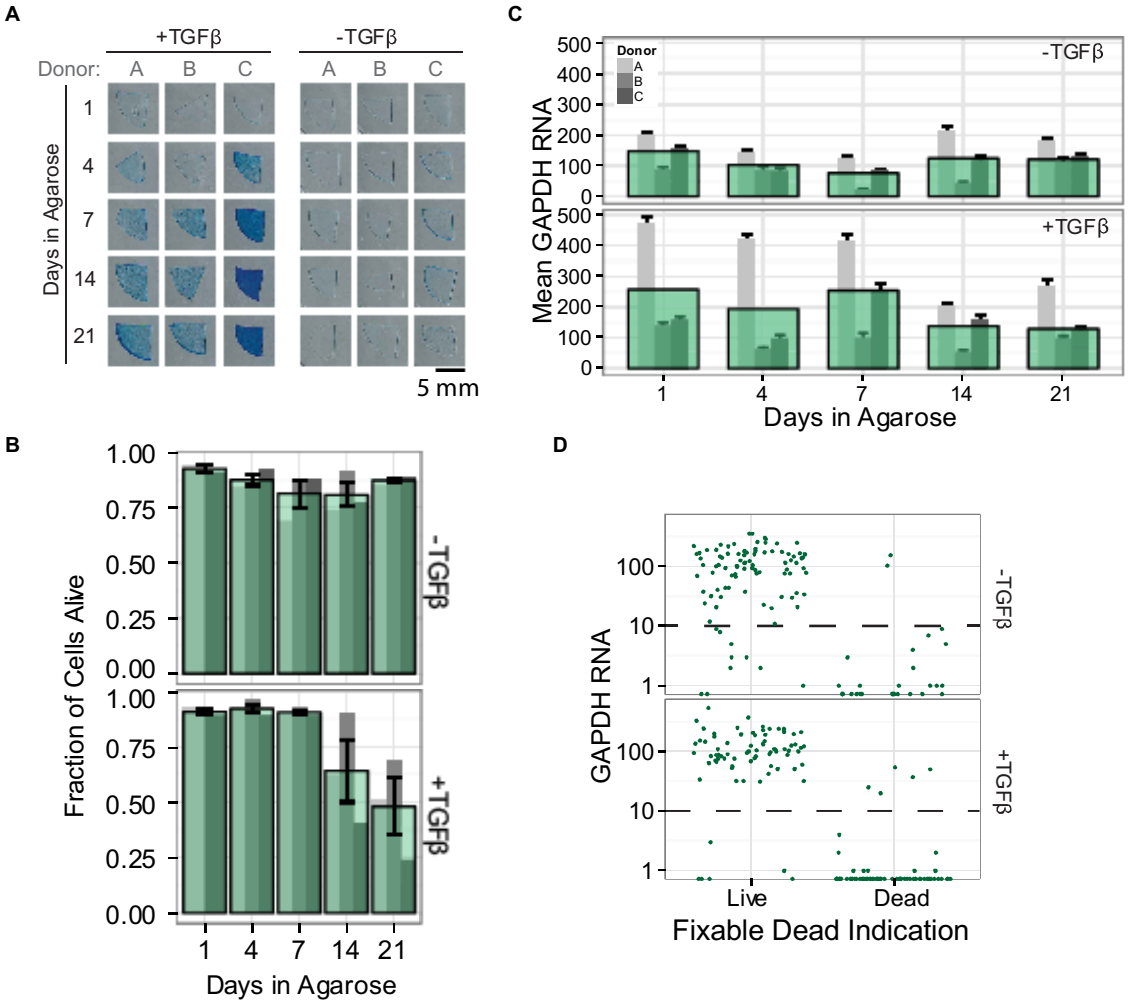
My splicing results challenge the assumption that splicing is happening immediately post-transcriptionally, and suggest that transcripts dwell at the site of transcription after transcription is complete. The timing of splicing is well studied, and the global percentages of co-transcriptional splicing found by others ((Brugiolo et al., 2013), 16-55% globally post-transcriptional) are in agreement with the percent of mRNA that I see as distally post-transcriptionally spliced (0-47%). However, my findings with expansion microscopy also suggest that there is a great deal of post-transcriptional splicing proximal to the site of

transcription. This is in agreement with more stringent fractionation and sequencing methods (Drexler et al., 2019), but in contrast to many traditional fractionation methods (Carrillo Oesterreich et al., 2016; Tilgner et al., 2012).

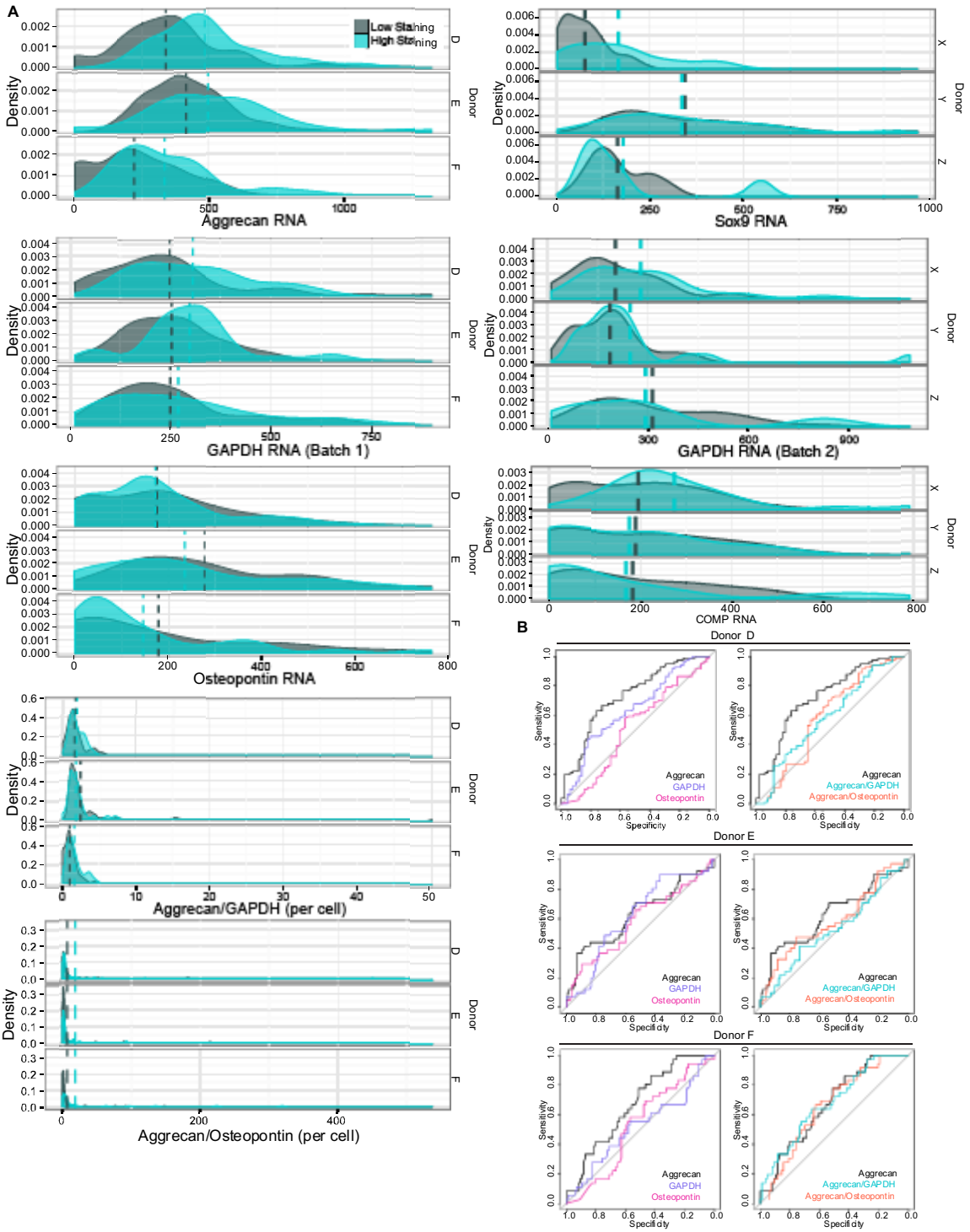
In the future, I would continue this work by continuing to use a combination of expansion microscopy and RNA FISH to confirm the dwell time and lack of tether observations for other genes of interest. In addition, it seems likely that at least some of the regulation of timing of splicing of these introns is due to sequence specificity, but I was unable to determine what the specific source of this sequence specificity was (data not shown), It would be of interest to take sequences, compare to mean dispersal or transcription site size, and analyze which features can explain the greatest diversity in splicing dynamics. Similarly to published work (Drexler et al., 2019), I found no sequence determinants that could explain my observations of post-transcriptionality, but it is likely that there is some currently unexplored aspect of sequence specificity that could explain these differences. This could potentially also be illuminated by experiments such as pulldown of the RNAs of interest (a comparison of highly dispersed and highly undispersed) followed by a proteomics based quantification of their respective RNA binding proteins, which may illuminate which factors are responsible for the differences in splicing rates, if any. It is possible that this difference in splicing rates could be also be based on the presence of alternative splice sites within the introns of interest, potentially “confusing” the spliceosome, and forcing the spliceosome to take longer to identify the canonical splice sites. This would also be illuminated by further exploration into the sequence specific differences between the multiple introns that exhibit different dispersal patterns.

Taken together, my data suggest that cell-to-cell heterogeneity can be due to many factors, including variability in transcription and splicing in single cells.

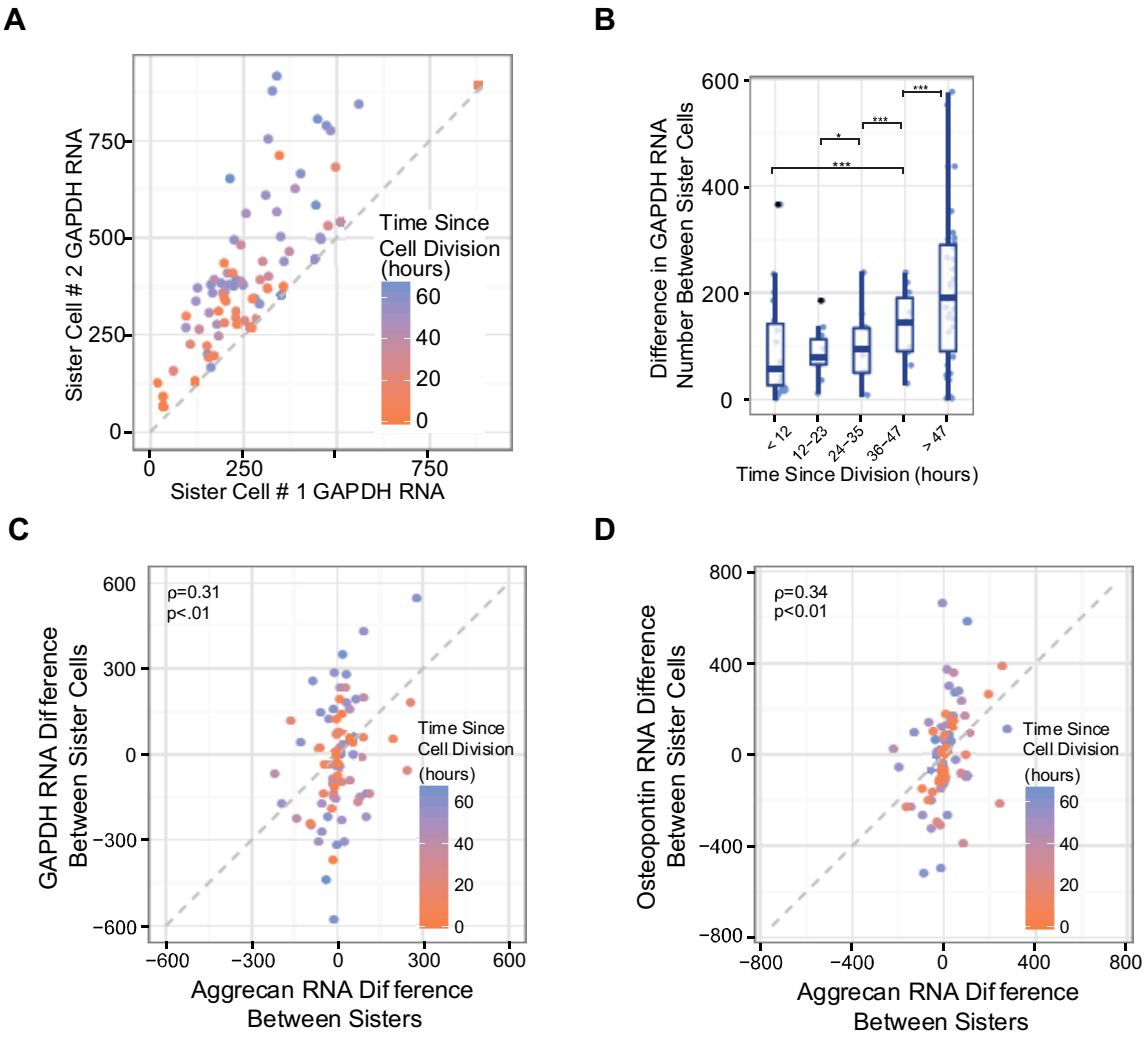
Supplementary Figures



Supplementary Figure 1: Matrix production, GAPDH copy number, and viability of MSCs in 3D culture. a) Alcian blue staining for sulfated proteoglycans for MSCs in 3D culture, for all donors and time points, with and without TGFβ induction of chondrogenesis. Donor B is also shown in Figure 1. Scale bar = 5 mm. b) Mean GAPDH RNA counts and c) cell viability over 21 days in 3D culture. Narrow bars represent the mean within an individual donor; overlaid bars represent the mean across donors. Error bars indicate standard error (n = 24-128 cells per donor and condition). RNA count means compared by t-tests with Satterthwaite approximation and simulated adjustment for multiple comparisons. See Supplementary Table 4 for all statistical comparisons. d) Simultaneous RNA FISH and fixable dead staining established a threshold of GAPDH>10 to differentiate live cells from dead cells for further analysis.
n = 85 cells for TGFβ-, 75 cells for TGFβ+.

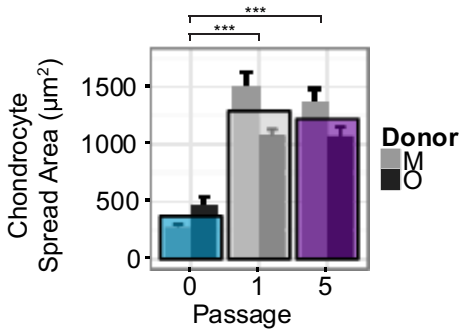

A
Low Staining
High Staining
Density
Donor
Aggrecan RNA
Sox9 RNA
GAPDH RNA (Batch 1)
GAPDH RNA (Batch 2)
COMP RNA
Osteopontin RNA
Aggrecan/GAPDH (per cell)
Aggrecan/Osteopontin (per cell)
B
Donor D
Donor E
Donor F
Sensitivity
Specificity
Aggrecan
GAPDH
Osteopontin
Aggrecan/GAPDH
Aggrecan/Osteopontin

Supplementary Figure 2: Matrix staining intensity versus mRNA copy number and ROC curves for individual donors, markers, and marker ratios. a) Distribution of aggrecan, GAPDH, osteopontin, aggrecan/GAPDH, and aggrecan/osteopontin gene expression within high- and low-performing MSC populations; separated by donor. Dashed lines represent the mean for each condition. For aggrecan/osteopontin, only cells that had at least one osteopontin (>95% of cells per donor) mRNA were included in the graph and mean statistics calculation. b) Receiver operating characteristic curves using individual gene expression and gene expression ratios to distinguish between high- and low-performing MSCs, separated by donor. Cells/donor: D: 132, E: 153, F: 122, X: 57, Y: 42, Z: 47.

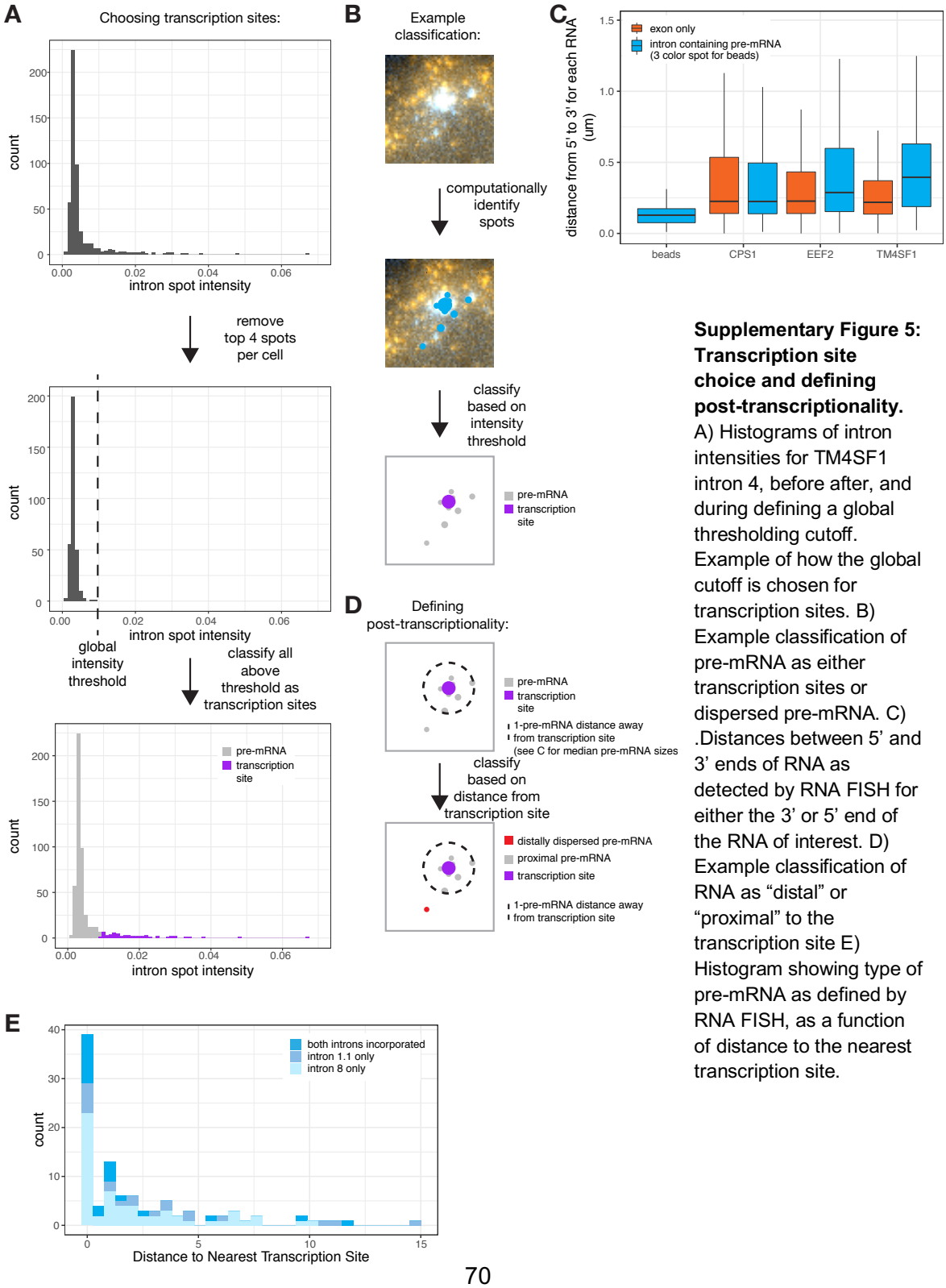


Supplementary Figure 3: Heritability of marker copy number through cell division.
a-b) Divergence in gene expression between sister cells as a function of time since their last division. c) Divergence in GAPDH vs aggrecan and d) osteopontin vs aggrecan between sister cells as a function of time since their last division.

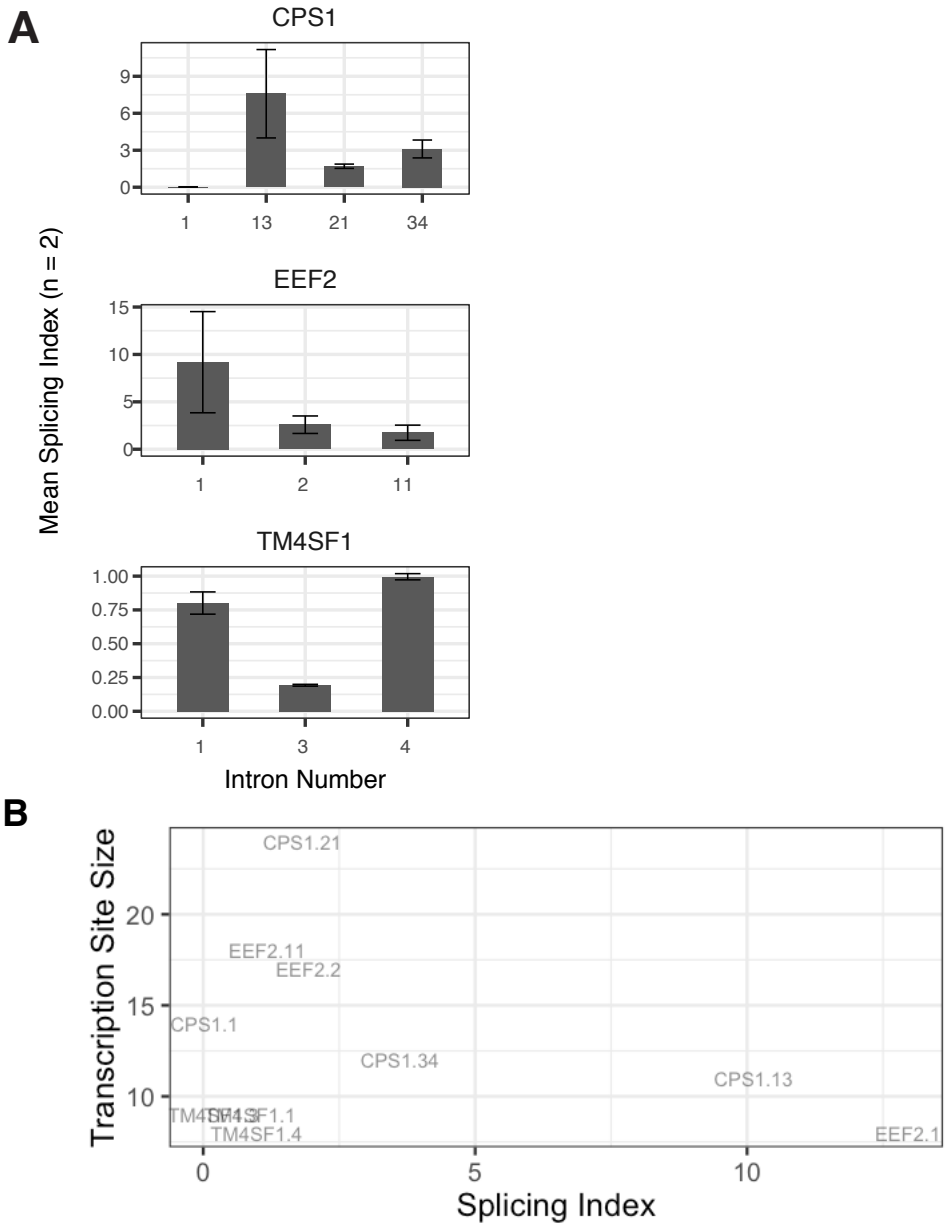
Box hinges denote the first and third quartiles. Whiskers extend from the hinges to the most extreme data points within (1.5 * interquartile range) of the hinges. Means compared by t-tests with Satterthwaite approximation and simulated adjustment for multiple comparisons, *** indicates p<0.001,* p<0.05. See Supplementary Table 6 for all statistical comparisons. n = 81 sister cell pairs.



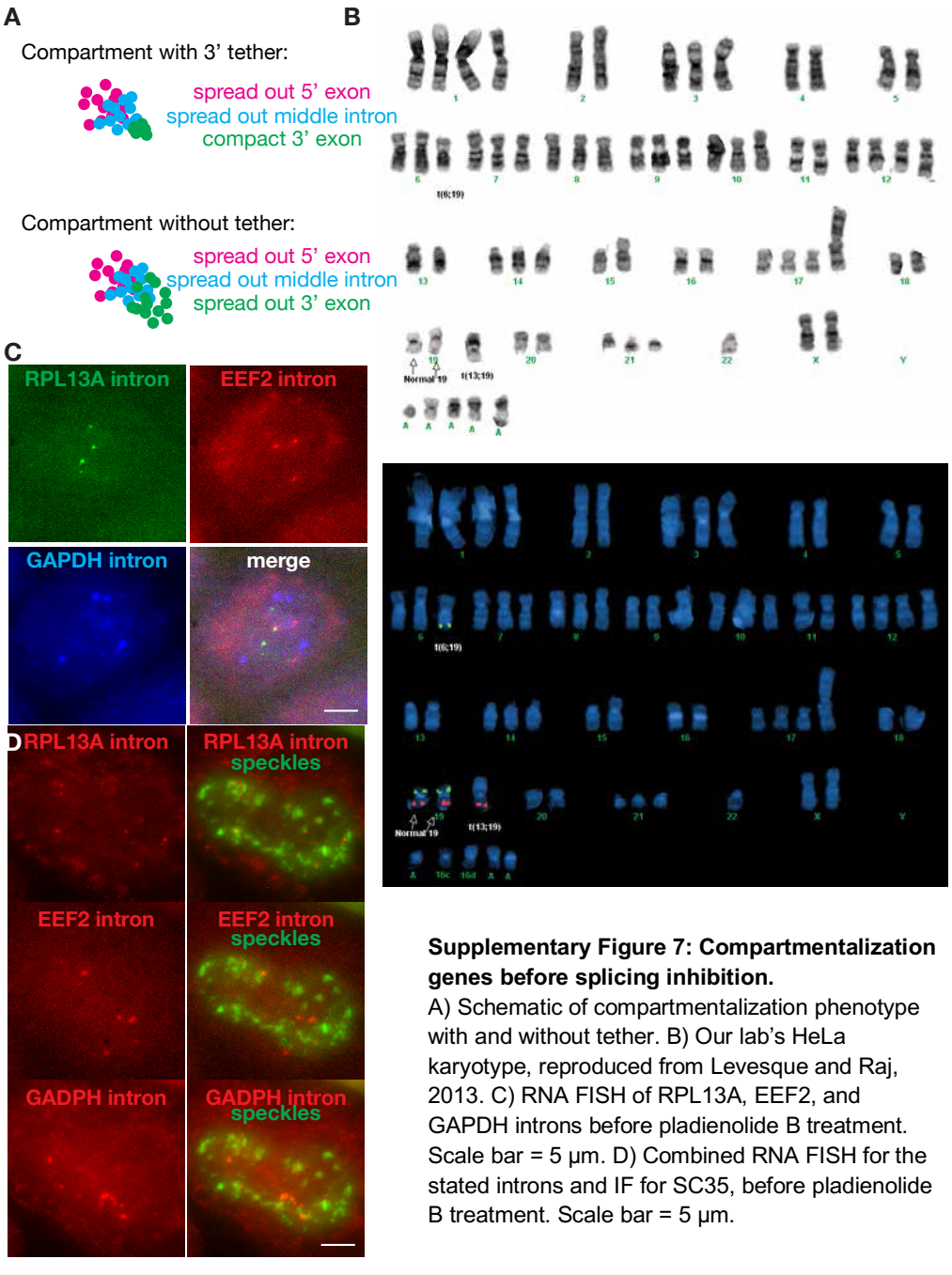
Supplementary Figure 4: Chondrocyte morphology with passage number.
Chondrocyte spread area with increasing passage number during de-differentiation (n = 25-27 cells per donor per condition). Narrow bars represent the mean within an individual donor; overlaid bars represent the mean across donors. Error bars indicate standard error. Compared via one-way ANOVA with Tukey post-hoc test, *** indicates $p<0.01$.



Supplementary Figure 5: Transcription site choice and defining post-transcriptionality. A) Histograms of intron intensities for TM4SF1 intron 4, before after, and during defining a global thresholding cutoff. Example of how the global cutoff is chosen for transcription sites. B) Example classification of pre-mRNA as either transcription sites or dispersed pre-mRNA. C) .Distances between 5' and 3' ends of RNA as detected by RNA FISH for either the 3' or 5' end of the RNA of interest. D) Example classification of RNA as "distal" or "proximal" to the transcription site E) Histogram showing type of pre-mRNA as defined by RNA FISH, as a function of distance to the nearest transcription site.



Supplementary Figure 6: Splicing index varies on a per intron basis. A) Splicing index of each intron for which we obtained RNA FISH measurements. Error bars represent mean +/- sd. n = 2 B) Comparison of transcription site size (as assessed by RNA FISH) and splicing index, as assessed by sequencing.



Supplementary Figure 7: Compartmentalization genes before splicing inhibition.
A) Schematic of compartmentalization phenotype with and without tether. B) Our lab's HeLa karyotype, reproduced from Levesque and Raj, 2013. C) RNA FISH of RPL13A, EEF2, and GAPDH introns before pladienolide B treatment. Scale bar = 5 μm. D) Combined RNA FISH for the stated introns and IF for SC35, before pladienolide B treatment. Scale bar = 5 μm.

Supplementary Tables

Supplementary Table 1: Correlation in Expression of Different Marker Genes in Differentiating MSCs			
Timepoint/Gene Pair	**r^2**	**Pearson's correlation coefficient**	**p-value**
Day 1 aggrecan - osteopontin	0.24	0.49	1.0e-07
Day 21 aggrecan - osteopontin	.11	.34	.0019
Day 1 aggrecan-LPL	7.5e-5	.0087	.92
Day 21 aggrecan-LPL	.0022	.047	.67
Day 1 osteopontin-LPL	1.9e-5	.0044	.96
Day 21 osteopontin-LPL	.04	.21	.059

Supplementary Table 2: Coefficient of Variation of RNA Count in Small MSC Colonies				
Colony	**A**	**B**	**C**	**D**
Aggrecan CV	1.71	1.34	1.07	0.91
GAPDH CV	0.32	0.35	0.31	0.20

Supplementary Table 3 P-values for Figure 1E: Aggrecan RNA Count, MSCs with Differentiation	
Comparison	**Adjusted P-value**
-TGFB vs +TGFB, Day 01	0.0067
-TGFB vs +TGFB, Day 04	0.0043
-TGFB vs +TGFB, Day 07	0.0020
-TGFB vs +TGFB, Day 14	0.0017
-TGFB vs +TGFB, Day 21	0.0031
+TGFB, Day 1 vs Day 4	0.0006
+TGFB, Day 1 vs Day 7	0.0002
+TGFB, Day 1 vs Day 14	0.0047
+TGFB, Day 1 vs Day 21	0.0002
+TGFB, Day 4 vs Day 7	0.0004
+TGFB, Day 4 vs Day 14	0.0021
+TGFB, Day 4 vs Day 21	0.0002
+TGFB, Day 7 vs Day 14	0.0002
+TGFB, Day 7 vs Day 21	<.0001
+TGFB, Day 14 vs Day 21	0.0002

Supplementary Table 4 P-values for Figure S1B: GAPDH RNA Count , MSCs with Differentiation	
Comparison	**Adjusted P-value**
-TGFB vs +TGFB, Day 01	0.3572
-TGFB vs +TGFB, Day 04	0.0827
-TGFB vs +TGFB, Day 07	0.0361
-TGFB vs +TGFB, Day 14	0.9479
-TGFB vs +TGFB, Day 21	0.9980
+TGFB, Day 1 vs Day 4	0.0022
+TGFB, Day 1 vs Day 7	0.4637
+TGFB, Day 1 vs Day 14	<.0001
+TGFB, Day 1 vs Day 21	0.0002
+TGFB, Day 4 vs Day 7	0.0019
+TGFB, Day 4 vs Day 14	0.0002
+TGFB, Day 4 vs Day 21	0.0009
+TGFB, Day 7 vs Day 14	<.0001
+TGFB, Day 7 vs Day 21	0.0002
+TGFB, Day 14 vs Day 21	0.0016

Supplementary Table 5
P-values for Figure S3:
Aggrecan Difference Between Sister Cells

Comparison		Adjusted P-value
< 12 hrs	[12-23 hrs]	<.0001
< 12 hrs	[24-35 hrs]	<.0001
< 12 hrs	[36-47 hrs]	<.0001
< 12 hrs	> 47 hrs	<.0001
[12-23 hrs]	[24-35 hrs]	0.8759
[12-23 hrs]	[36-47 hrs]	0.7564
[12-23 hrs]	> 47 hrs	0.9997
[24-35 hrs]	[36-47 hrs]	<.0001
[24-35 hrs]	> 47 hrs	<.0001
[36-47 hrs]	> 47 hrs	<.0001

Supplementary Table 6
P-values for Figure S3:
GAPDH Difference Between Sister Cells

Comparison		Adjusted P-value
< 12 hrs	[12-23 hrs]	0.1649
< 12 hrs	[24-35 hrs]	0.6435
< 12 hrs	[36-47 hrs]	<.0001
< 12 hrs	> 47 hrs	<.0001
[12-23 hrs]	[24-35 hrs]	0.0173
[12-23 hrs]	[36-47 hrs]	<.0001
[12-23 hrs]	> 47 hrs	<.0001
[24-35 hrs]	[36-47 hrs]	<.0001
[24-35 hrs]	> 47 hrs	<.0001
[36-47 hrs]	> 47 hrs	<.0001

Supplementary Table7
P-values for Figure 4B-D
Chondrocyte Gene Expression During Dedifferentiation

Passage Comparison		Aggrecan/GAPDH (Pooled Population) Adjusted P-value	Aggrecan per Cell Adjusted P-value	GAPDH per Cell Adjusted P-value
0	1	0.0079	0.2108	0.0006
0	3	0.0004	0.9342	0.0006
0	5	0.0002	0.9999	0.0008
0	7	0.0002	1.0000	0.0008
0	9	0.0011	0.9984	0.0025
1	3	0.7036	0.6621	1.0000
1	5	0.5040	0.2891	0.9945
1	7	0.4593	0.1658	0.9929
1	9	0.9377	0.3623	0.7026
3	5	0.9993	0.9783	0.9981
3	7	0.9982	0.8850	0.9972
3	9	0.9944	0.9934	0.7784
5	7	1.0000	0.9988	1.0000
5	9	0.9510	0.9999	0.9343
7	9	0.9309	0.9941	0.9431

Supplementary Table 8
P-values for Figure 4H-J
Chondrocyte RNA Count During Redifferentiation

Comparison	Aggrecan per Cell Adjusted P-value	GAPDH per Cell Adjusted P-value	Aggrecan/GAPDH per Cell Adjusted P-value
Day 1, P0 vs P5	0.0016	0.0406	0.9983
P0, Day 1 vs 14	0.0897	0.0006	0.0176
Day 14, P0 vs P5	0.0002	0.6714	0.0588
P5, Day 1 vs 14	<.0001	<.0001	0.1982

Supplementary Table 8: RNA-FISH probe sequences.			
Aggrecan	**GAPDH**	**Osteopontin**	**Lipoprotein Lipase**
gtccttgtctccatagcaac	agaattaaaagcagccctgg	cccaagaggcagaagcaaat	ggagtcctgagagcaaattt
acacgtcataggtttcgttg	tcattgatggcgacgatgtc	gtttaactggaagggcggag	caggaatgaggtggcaagtg
cttggaaggtgaacttctcc	agaccatgtagtgaaggtca	tacagcatctgggtatttgt	tgctgtggttgaagtgacag
tttctgggatgtccacaaag	ttgactgtgccgttgaactt	aagtctgcttctgagatggg	caactctcatacattcctgt
cccttacttcagggacaaac	ctttccattgatgacgagct	gacacagaattctgtggtgc	gacactggataatgctgctg
actgatgtcctctactccag	gatctcgctcctggaagatg	gggtattttgtttgttgtca	ataaacttggccacatcctg
gcagagatttctggaccttc	tctccatggtagtgaagacc	ggctttcattggacttactt	agttaaattcatccgccatc
actgatgtcctctactccag	agagatgatgaccctcttgg	gtcatctagatcgtctgttt	atcccaagagatgcacattg
aaggtcctctactccagaag	ttgttatacttctcgtggtt	cgtcggagtcattagagttg	ttcttattggtcagacttcc
agaaggaagtccactaaggt	gaggcattgctgacaatctt	aatggtcagggtcatcagtg	catactcgaagttaggtcca
agtctgctgagatcctctac	ccaaagtggtcatggatgac	tcagaatggtgagactcgtc	caggagaaaggcgacttgga
tctactccagaagcagagac	aagcagggatgatattctgg	gcgattgttggaatatcagt	ccctggtgaatgtgtgtaaa
actgatgtcctctactccag	ttgagctcagggatgacctt	ccgtagggataaacggagtg	ttctggattccgatacttcg
caccagaaggaagaacactg	gtcagatccacaacagacac	tcttgacttcagtccgtaag	cgttagggtaaatgtccaca
actgatgtcctctactccag	catacttggcaggtttctcc	ttagatcggcggaacttctt	caatgttacatcctggttgg
actgatgtcctctactccag	gtcctcagtgtagcctagaa	tgtggcatctggactctgaa	ctctgcaatcacacggagag
cctccagaaggaagtctact	tgagtgtcgctgttgaagtc	ctgactcgtcttcttaggtg	taggccttacttggattttc
agaaggaagtccactaaggt	aatgagcttgacaaagtggt	tggcgtgagttctttggaaa	ggcagagacctttctcaaag
gaaggtattctgccaaggtc	ttgctgtagccgaattcatt	gaatgcttgttcttatcctt	tgaccttgttgatctcgtag
gtccactgaaatcaggaagg	tactccttggaggccatgtg	tttcctgactctcaatcaga	tacatcttgctgcttctttt
caaagagtccagaggtttcc		aattcttggctgagtttgga	gcatctgagaacgagtcttc
ccactaatgtcaggaaaccc		atgatctaggtctagcttgt	cctggttggtgtatgtatta
gtcacgccagatatttctcc		tttcaggtgtttgtcttctt	cagtgccatacagagagatc
gatttctggttgtccagagg		attgacctcagaagaggcac	aaaagggatgttctcgctct
caattgctccagaagacagt			ttgtggaaacttcaggcagg
aagttccagaatgcctgaag			aatatccacctccgtgtaaa
ccaaagatcccaaatggtct			ggaccagctgaagtaggaat
accactggattcaaaaagct			cacctttttttgagtctctc
tgggactgatgacacttcta			gataagacattttctcccgg
cagatcagcttcatggaagg			tatcacaggtgactttcctt
ggattcgatttctagacgcg			actttgtcatggcatttcaca
attgatctcgtatcggtcct			cagccagactttctattcag

Appendix 1: Materials and Methods

Chapter 2 Methods

Cell isolation and expansion

MSCs were isolated from the tibial and femoral bone marrow of juvenile bovine cows (3–6 months, Research 87, Boylston, MA) and expanded in a basal media consisting of high-glucose DMEM with 10% FBS and 1 × antibiotic–antimycotic. After the initial plating reached ~80% confluence, cells were passaged at a ratio of 1:3 before use in experiments. For single-cell-derived colonies, bovine MSCs were isolated as described above and seeded sparsely onto glass coverslips. Individual colonies were allowed to expand for 3 days in basal media, followed by 4 days in chondrogenic induction media before fixation. All cells in each colony were manually located and imaged as described below. Chondrocytes were isolated from articular cartilage from the trochlear groove of juvenile bovine knees. Cartilage was digested in basal media supplemented with type II collagenase (0.5 mg ml−1, Sigma-Aldrich) for up to 18 h. Isolated cells were filtered, washed and plated in basal media. To improve cell yield for chondrocyte re-differentiation studies, cartilage was also digested in basal media with pronase (2.5 mg ml−1, Calbiochem) for 1 h before collagenase digestion. For all studies, chondrocytes were expanded in basal media and passaged 1:10 when plates reached ~80% confluence. All bovine cells were derived from animals used in the food production industry, and so no institutional approvals were required.

Cell encapsulation

For 3D culture, MSCs (passage 2) or chondrocytes (passage 0 and passage 5) were encapsulated in 2% agarose microgels at a density of two million cells per ml. Molten 4%

w/v agarose (type VII, Sigma, 44 °C) was mixed 1:1 with cells suspended in media and pipetted into small drops in a well plate. Round coverslips were placed on top of the molten drops to spread the mixture before the gel solidified, resulting in the formation of uniform microgels that were 10–12 mm in diameter (depending on coverslip diameter) and ~400 μm thick. Coverslips were removed from the microgels before culture. Microgels were supplied with fresh media every 3 days and 24 h before collection. MSC microgels were maintained in a chemically defined media consisting of high-glucose DMEM supplemented with 1 × antibiotic–antimycotic, 40 ng ml−1 dexamethasone, 50 μg ml−1 ascorbate 2-phosphate, 40 μgml−1 l-proline, 100 μg ml−1 sodium pyruvate, 1.25 mg ml−1 bovine serum albumin, 5.35 μg ml−1 linoleic acid and 1 × insulin–transferrin–selenous acid premix (Corning CB-40350), either with or without 10 ng ml−1 TGFβ3 (R&D Systems) (Mauck et al., 2006). Chondrocyte microgels were cultured in basal media (high-glucose DMEM+10% FBS+1 × antibiotic–antimycotic) supplemented with 50 μg ml−1 ascorbate 2-phosphate. At defined time points, gels were fixed for 30 min in paraformaldehyde (PFA) and stored in 70% ethanol at 4 °C.

Cell viability in gels was measured using the LIVE/DEAD Cell Viability Assay Kit (Molecular Probes L-3224). A custom Matlab script quantified the number of live (calcein-AM positive) and dead (ethidium-homodiner-1 positive) cells in three 4 × fields of view per microgel. To measure viability in conjunction with RNA FISH, a fixable, amine-binding green fluorescent dead cell stain (Molecular Probes L-23101) was employed. For fixable dead staining, microgels were washed with PBS, stained for 30 min in a 1:5,000 dilution in PBS, washed with PBS again, and then fixed in PFA before RNA FISH analysis, as described below.

Chondrogenic pellet culture and biochemical content

Clonally derived passage 2 MSCs were formed into cell-rich pellets via centrifugation (200,000 cells per pellet) and cultured in chondrogenic induction media with TGFβ for 21 days (Farrell et al., 2015). Pellets were papain digested and biochemically assayed for glycosaminoglycan and DNA content using 1,9-dimethylmethylene blue and Picogreen (Molecular Probes, Eugene, OR) assays, respectively (Farrell et al., 2015).

Live cell imaging and tracking

To quantify mRNA levels as a function of the time history of division, passage 2 MSCs were seeded into two-well LabTek chambered coverglass dishes (Fisher Scientific) and cultured in chondrogenic induction media with TGFβ for 4 days. Seeded cells were supplied with fresh media every 3 days and 24 h before fixation. Over the last 3 days of culture, live cells were imaged using a Nikon Ti-E microscope with a custom environmental chamber. Transmitted light images were automatically acquired every 30 min over a period of 70 h using a × 10 air objective over a 289-image grid in each well of the two-well coverglass. Cell division was tracked manually using ImageJ, and matched to the corresponding RNA FISH quantification that followed.

RNA fluorescence in situ hybridization and imaging

Single-molecule RNA FISH was performed on samples (Raj et al., 2008). Microgels and monolayer cells were fixed in PFA and permeabilized with 70% ethanol before in situ hybridization was performed using the specified pools of oligonucleotides. Monolayer and microgel samples were simultaneously co-stained with oligonucleotide probes for

osteopontin labelled with Cy3, LPL labelled with Alexa 594, aggrecan labelled with Atto 647 N and GAPDH labelled with Atto700 (Stellaris oligonucleotides, Biosearch Technologies). See Supplementary Table 9 for a complete list of sequences of oligonucleotide probes used in this chapter 2. Subsequently, samples were washed with 2 × saline sodium citrate buffer (SSC) with 10% formamide (Ambion), and then 2 × SSC supplemented with DAPI (Molecular Probes D3571) to stain the cell nuclei. Monolayer cells cultured in coverglass chambers were submerged in 2 × SSC for imaging. Microgels were mounted in 2 × SSC and compressed between a coverglass and slide for imaging. Cells in the microgel and small colonies were imaged using a Leica DMI600B automated widefield fluorescence microscope equipped with a × 100 Plan Apo objective, a Pixis 1024BR cooled CCD (charge-coupled device) camera, a Prior Lumen 220 light source, and filter sets specific for each fluorophore. Images in each fluorescence channel were taken as a series of optical z-sections (0.5–0.7 microns per section) spanning the vertical extent of each cell. To prevent differences in viability between conditions from confounding interpretation of single-cell gene expression, the fixable dead cell stain was used to establish a GAPDH copy number of >10 mRNA as a threshold to identify live cells for inclusion in further analysis (Supplementary Fig. 1D). When this FISH analysis was applied to live-imaged cells, single plane scans were performed using a Nikon Ti-E microscope with a × 63 Plan Apo objective.

Quantification of copy number from RNA FISH images

On collecting images of RNA FISH samples, cell boundaries were manually identified and RNA spots were counted and localized using custom software written in MATLAB (Raj et al., 2008). For spot counting in FISH images from live cell tracking, each cell was tracked

through the acquired time series, and sister cells manually matched, with care taken to note the time since last division.

Quantification of extracellular matrix deposition

Extracellular aggrecan protein content was quantified by immunostaining. Briefly, after the final wash stages of the FISH protocol, samples were incubated with primary antibody (Abcam ab3778, 1:50 in PBS) at 4 °C overnight, washed for 30 min in PBS, incubated with Alexa 488 secondary antibody (Invitrogen, 1:200 in PBS) at room temperature for 1 h, washed with PBS for 30 min and then mounted for imaging. For immunofluorescence images, a scorer blinded to the RNA FISH images examined the DAPI, GFP (aggrecan core protein) and transmitted light images to classify cells with and without extracellular aggrecan core protein staining.

Receiver operating characteristic curves were constructed and analysed using the pROC package in R (Robin et al., 2011). Matrix deposition (high versus low) was used as the binary outcome, and sensitivity and specificity were calculated for possible thresholds of RNA copy number, a linear combination of RNA counts and RNA ratios. To construct the linear combination, the data sets corresponding to each batch (Batch 1: Donors D-F, assayed for aggrecan, osteopontin, LPL, GAPDH; Batch 2: Donors X-Z, assayed for COMP, Sox9 and GAPDH) were randomly split in half to create training and test data sets to be used for model construction and evaluation, respectively. Logistic regression was performed using glm in R, and non-significant terms were dropped. For batch 1, the final model was established as (equation (1)):

$$\ln\left(\frac{\widehat{\pi}_i}{1-\widehat{\pi}_i}\right) = \beta_{\mathrm{ACAN}}\left(\mathrm{aggrecan}_i\right) - \beta_{\mathrm{OPN}}\left(\mathrm{osteopontin}_i\right) - 1.52$$

the estimated probability of the i-th cell having high matrix staining, was a function of the cell's aggrecan and osteopontin expression. Aggrecan associated positively ($\beta_{ACAN}=0.003$, $P<0.001$), while osteopontin associated negatively ($\beta_{OPN}=-0.001$, $P=0.05$); other markers were not significantly associated with matrix deposition. For batch 2, the intercept was the only significant term and the model was not further analysed. Having established this model on the training data set, its predictive performance was evaluated by constructing an ROC curve of the model applied to the test data set.

RNA sequencing

Poly-adenylated RNA from passage 1 clonal MSCs populations were isolated from monolayer culture in basal media. The Qiagen miRNeasy kit was used for RNA isolation, the NEB Next Poly(A) mRNA Magnetic Isolation Module was used for selection of poly-adenylated transcripts and NEB Next Ultra Library Preparation Kit for Illumina was used for library preparation. Each sample was sequenced with 50-bp single-end reads on an Illumina HiSeq and to a depth of 15–25 M reads. Reads were aligned to bosTau7 using STAR69. Reads per gene were quantified using HTSeq and a RefSeq bosTau7 from annotation release 103 (Anders et al., 2015; Dobin et al., 2013). FPKM (fragments per kilobase of transcript per million mapped reads) for each gene was calculated using R.

Cell volume and area measurements

Chondrocyte area was measured in ImageJ by manually tracing images of phalloidin-stained cells sparsely plated onto glass coverslips. Chondrocyte suspended cell volumes were computed from the cell radii measured by an automatic cell counter (Nexcelom Cellometer) for chondrocytes in solution during passaging (immediately following trypsinization).

Alcian blue staining

Microgels were removed from 70% ethanol and equilibrated in 3% acetic acid for 30 min at room temperature. Gels were then transferred to Alcian blue solution (pH 1.0, Rowley Biochemical) for 30 min, washed three times in acid alcohol (1% hydrochloric acid in 70% ethanol) for 30 min and then washed in PBS for 30 min before imaging. For macroscopic images, gels were photographed using a Ricoh photocopier and digital camera. For microscopic images, microgels were mounted in PBS and compressed between a coverglass and slide. Images were taken at × 10 using a Nikon Eclipse Ni-E motorized upright microscope.

Statistical comparisons

To compare mean single-cell RNA counts, a generalized linear mixed model with a log-link function and by-donor random intercepts was constructed. For MSC RNA counts, media condition, and culture duration (with interaction term) were considered fixed effects, and an additional by-donor random slope effect was associated with media (−TGFβ versus +TGFβ). For MSC RNA divergence, time-since-division was considered a fixed effect. For chondrocyte RNA counts during de-differentiation in monolayer, passage was considered

a fixed effect and was also associated with a by-donor random slope. For chondrocyte RNA counts during re-differentiation, passage and culture duration (with interaction term) were considered fixed effects, and an additional by-donor random slope effect was associated with passage. In each model, estimated means were compared using Satterthwaite-based t-distributions with simulated adjustment for multiple comparisons (SAS Studio 3.3). Pooled chondrocyte aggrecan/GAPDH expression data were compared using a one-way analysis of variance with Tukey's post hoc test. Chondrocyte area and volume data were pooled across donors and compared using a one-way analysis of variance with Tukey's post hoc tests. Sample size was chosen based on previous experience with these assays. Details of all statistical comparisons are provided in the Supplementary Tables.

Chapter 3 Methods

Cell culture, splicing inhibition, and FKBP5 induction

HeLa (kind gift of the lab of Dr. Phillip Sharp, MIT) and A549 (human lung carcinoma, A549, ATCC CCL-185) cells were cultured in DMEM (Gibco) supplemented with 1X antibiotics and 10% fetal bovine serum (FBS, Fisher). Splicing inhibition was accomplished by treating HeLa cells with 1uM Pladienolide B (Tocris Biosciences, 6070500U) for 4 hours, as described by (Nojima et al., 2015). HeLa cells were then fixed and used for RNA FISH as described below. FKBP5 was induced by treating A549 cells with 25nM dexamethasone (Sigma, D2915) for the specified lengths of time. A549 cells were then fixed and used for RNA FISH as described below.

RNA fluorescence in situ hybridization and expansion microscopy

Single-molecule RNA FISH was performed on samples as described previously(Raj et al., 2006). Cells were fixed in formaldehyde and permeabilized with 70% ethanol before in situ hybridization was performed using the probes described. Samples were simultaneously co-stained with probes for the exon of gene of interest (labelled in cy3), two introns of the gene of interest (labelled in alexa594 or atto647N), and cyclin mRNA (labelled in either atto700 or atto647N) (Stellaris oligonucleotides, Biosearch Technologies). Samples were then washed with 2 X saline sodium citrate buffer (SSC) with 10% formamide (Ambion), and then 2XSSC supplemented with DAPI (Molecular Probes D3571) to stain the cell nuclei. Cells were submerged in 2XSSC with DAPI for imaging. For combined expansion microscopy and RNA FISH, expansion microscopy was performed as described by Chen *et al.* (Chen et al., 2016). Cells were fixed and permeabilized as described above, before a poly-acrylamide gel was polymerized on top

of the sample for 1 hour at 37°C. Cells were digested with proteinase K overnight at room temperature, then RNA FISH was performed as described above. Samples were then expanded in nuclease free water for 2 hours at room temperature, then expanded for an additional 2 hours at room temperature in nuclease free water with DAPI. Samples were submerged in nuclease free water with DAPI for imaging.

Immunofluorescence

Staining for SC35 and polymerase II were performed with antibodies against SC35 (abcam ab11826, 1:200) and phospho S2 polymerase II (Active Motif, 61083, 1:200), respectively. Briefly, staining was performed on cells fixed and permeabilized as described above for RNA FISH. Primary antibody hybridization was carried out in 1XPBS overnight at 4°C. Samples were then washed with 1XPBS and incubated with secondary antibody for 1 hour in 1XPBS at room temperature. Samples were then washed with 1XPBS and RNA FISH was performed as described above.

Imaging

Cells were imaged using a Leica DMI600B automated widefield fluorescence microscope equipped with a X100 Plan Apo objective, a Pixis 1024BR cooled CCD (charge-coupled device) camera, a Prior Lumen 220 light source, and filter sets specific for each fluorophore. Images in each fluorescence channel were taken as a series of optical z-sections (0.3 microns per section).

RNA FISH quantification

RNA FISH was quantified as described previously (Raj et al., 2006). Briefly, cells were manually segmented, a gaussian filter was applied to all spots, signal was distinguished from noise through semi-automated thresholding, each called spot was further fit to a gaussian to get sub pixel resolution, and transcription sites were chosen based on a global brightness threshold (Supplementary Figure 5A). Data was processed to assess distances and graphed in R.

4sU labeled chromatin-associated RNA sequencing and splicing index analysis.

HeLa S3 cells (ATCC, CCL-2.2) were maintained in DMEM media containing 10% FBS, 100 U/ml penicillin, and 100 µg/ml streptomycin to 75% confluency. Cells were labeled in media containing 500 µM 4-thiouridine (4sU, Sigma, T4509) for 7.5 minutes. Plates were washed twice with 1X PBS and cells were lifted by scraping. Labeled cells were collected by centrifugation at 500 g for 2 minutes. To purify chromatin associated RNA, steps 8-21 were followed exactly as described in (Mayer and Churchman, 2016). In brief, nuclei were collected by lysing samples of 10M cells in 200 µl cytoplasmic lysis buffer (0.15% (vol/vol) NP-40 (Thermo Fisher Scientific, 28324), 10 mM Tris-HCl (pH 7.0), and 150 mM NaCl) for 2 min, layering over a 500 µl sucrose cushion (10 mM Tris-HCl (pH 7.0), 150 mM NaCl, 25% (wt/vol) sucrose), and centrifuging at 16,000 g for 10 minutes. The nuclei pellet was washed in 800 µl wash buffer (0.1% (vol/vol) Triton X-100, 1 mM EDTA, in 1X PBS) and collected by centrifuging at 1,150 g for 1 minute. Nuclei were resuspended in 200 µl glycerol buffer (20 mM Tris-HCl (pH 8.0), 75 mM NaCl, 0.5 mM EDTA, 50% (vol/vol) glycerol, 0.85 mM DTT), and mixed with 200 µl nuclei lysis buffer (1% (vol/vol) NP-40, 20 mM HEPES (pH 7.5), 300 mM NaCl, 1 M urea, 0.2 mM EDTA, 1 mM DTT) by pulse vortex

and incubated on ice for 2 minutes. The chromatin pellet was collected by centrifugation at 18,500 g for 2 minutes and resuspended in 1X PBS. All steps were performed at 4°C and buffers were prepared with 25 µM α-amanitin (Sigma, A2263), 0.05U/µl SUPERase.In (ThermoFisher Scientific, AM2694), and protease inhibitor mix (Roche, 11873580001).

Chromatin-associated RNA was extracted using Qiazol lysis reagent (Qiagen, 79306) following the manufacturer's instructions. 50 µg RNA per reaction was subjected to 4sU purification as described in (Dölken et al., 2008; Schwalb et al., 2016). In brief, labeled RNA (1 µg / 10 µl) was incubated with 10% biotinylation buffer (100mM Tris pH 7.5, 10mM EDTA) and 20% EZ-Link Biotin-HPDP (1 mg/mL resuspended in DMF, Thermo Fisher Scientific, 21341) for 1.5 hours 24°C in the dark and 800 rpm to mix. RNA was purified by shaking the sample with a 1:1 volume of chloroform/isoamylacohol (24:1), separating using a phase-lock tube at 16,000 g for 5 min, and performing isopropanol precipitation. Biotinylated RNA was separated using the µMACS streptavidin kit (Miltenyi Biotec, 130-074-101) by mixing with µMACS streptavidin beads at a 2:1 ratio by volume at 800 rpm and 24°C for 15 min. RNA-streptavidin beads mix was transferred to the µMACS column and washed with wash buffer (100 mM Tris pH 7.5, 10 mM EDTA, 1 M NaCl, 0.1% Tween 20) at 65°C and room temperature 3 times each. Selected RNA was eluted off the magnet using the reducing agent, DTT (0.1M), and purified using the miRNeasy micro kit (Qiagen, 217084) with on-column DNase I treatment (Qiagen, 79254). For the poly(A) depleted sample, the RNA was first concentrated using the RNA Clean and Concentrator kit (ZymoResearch, R1015). 10 µL Oligo(dT) Dynabeads (ThermoFisher, 61002) were washed in 10 µL Binding Buffer (20mM Tris.HCl pH 7.5, 1M LiCl, and 2mM EDTA). The sample was mixed with 10 µl binding buffer, heated to 65°C for 2 min, moved to ice, and mixed with 1 µl SUPERase.In (ThermoFisher Scientific, AM2694). The sample and beads

were mixed thoroughly and annealed by rotating continuously on a mixer for 5 minutes at room temperature. Poly(A) RNAs were collected on a magnet while depleted supernatant was removed and purified using the RNA Clean and Concentrator kit (ZymoResearch, R1015). Illumina sequencing libraries were prepared using the Ovation Universal RNA-seq System (NUGEN, 0343-32) with Universal Human rRNA strand selection reagent (NUGEN, S01859) following the manufacturer's instructions.

All samples were sequenced 2x80 on a NEXTseq 500 sequencer (Illumina, San Diego, CA, USA) in the Biopolymers Facility at Harvard Medical School. Paired-end reads were aligned to the ENSEMBLE GRCh38 (release-86) reference genome using STAR (v2.5.1a) (Dobin et al., 2013) with default parameters (except for readFilesCommand=cat, limitIObufferSize=200000000, limitBAMsortRAM=64000000000, outReadsUnmapped=Fastx, outSAMtype=BAM SortedByCoordinate, outSAMattributes=All, outFilterMultimapNmax=101, outSJfilterOverhangMin=3 1 1 1, outSJfilterDistToOtherSJmin=0 0 0 0, alignIntronMin=11, alignEndsType=EndToEnd). Splicing index calculations were determined by summing the number of spliced and unspliced read pairs that span exon junctions by at least 3 nucleotides and calculating the total spliced read pairs divided by the total unspliced read pairs for each gene; splicing index = 2 × spliced read pairs / (5'SS unspliced + 3'SS unspliced read pairs).

References

Aebi M, Weissman C. 1987. Precision and orderliness in splicing. *Trends Genet* **3**:102–107.

Alexander RD, Barrass JD, Dichtl B, Kos M, Obtulowicz T, Robert M-C, Koper M, Karkusiewicz I, Mariconti L, Tollervey D, Dichtl B, Kufel J, Bertrand E, Beggs JD. 2010a. RiboSys, a high-resolution, quantitative approach to measure the in vivo kinetics of pre-mRNA splicing and 3'-end processing in Saccharomyces cerevisiae. *RNA* **16**:2570–2580.

Alexander RD, Innocente SA, Barrass JD, Beggs JD. 2010b. Splicing-dependent RNA polymerase pausing in yeast. *Mol Cell* **40**:582–593.

Alford JW, Cole BJ. 2005. Cartilage restoration, part 1: basic science, historical perspective, patient evaluation, and treatment options. *Am J Sports Med* **33**:295–306.

Alpert T, Herzel L, Neugebauer KM. 2016. Perfect timing: splicing and transcription rates in living cells. *Wiley Interdiscip Rev RNA*. doi:10.1002/wrna.1401

Altschuler SJ, Wu LF. 2010. Cellular heterogeneity: do differences make a difference? *Cell* **141**:559–563.

Ameur A, Zaghlool A, Halvardson J, Wetterbom A, Gyllensten U, Cavelier L, Feuk L. 2011. Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nat Struct Mol Biol* **18**:1435–1440.

Anders S, Pyl PT, Huber W. 2015. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**:166–169.

Baralle FE, Giudice J. 2017. Alternative splicing as a regulator of development and tissue identity. *Nat Rev Mol Cell Biol* **18**:437–451.

Barrass JD, Reid JEA, Huang Y, Hector RD, Sanguinetti G, Beggs JD, Granneman S. 2015. Transcriptome-wide RNA processing kinetics revealed using extremely short 4tU labeling. *Genome Biol* **16**:282.

Barry F, Boynton RE, Liu B, Murphy JM. 2001. Chondrogenic differentiation of mesenchymal stem cells from bone marrow: differentiation-dependent gene expression of matrix components. *Exp Cell Res* **268**:189–200.

Beane OS, Darling EM. 2012. Isolation, characterization, and differentiation of stem cells for cartilage regeneration. *Ann Biomed Eng* **40**:2079–2097.

Bentley DL. 2014. Coupling mRNA processing with transcription in time and space. *Nat Rev Genet* **15**:163–175.

Benya PD, Shaffer JD. 1982. Dedifferentiated chondrocytes reexpress the differentiated collagen phenotype when cultured in agarose gels. *Cell* **30**:215–224.

Beyer AL, Bouton AH, Miller OL Jr. 1981. Correlation of hnRNP structure and nascent transcript cleavage. *Cell* **26**:155–165.

Bhatt DM, Pandya-Jones A, Tong A-J, Barozzi I, Lissner MM, Natoli G, Black DL, Smale ST. 2012. Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* **150**:279–290.

Bonaventure J, Kadhom N, Cohen-Solal L, Ng KH, Bourguignon J, Lasselin C, Freisinger P. 1994. Reexpression of cartilage-specific genes by dedifferentiated human articular chondrocytes cultured in alginate beads. *Exp Cell Res* **212**:97–104.

Bosnakovski D, Mizuno M, Kim G, Takagi S, Okumur M, Fujinag T. 2006. Gene expression profile of bovine bone marrow mesenchymal stem cell during spontaneous chondrogenic differentiation in pellet culture system. *Jpn J Vet Res*

53:127–139.
Brugiolo M, Herzel L, Neugebauer KM. 2013. Counting on co-transcriptional splicing. *F1000Prime Rep* **5**:9.
Budnik B, Levy E, Harmange G, Slavov N. 2018. SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol* **19**:161.
Cai L, Dalal CK, Elowitz MB. 2008. Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature* **455**:485–490.
Carballo CB, Nakagawa Y, Sekiya I, Rodeo SA. 2017. Basic Science of Articular Cartilage. *Clin Sports Med* **36**:413–425.
Carrillo Oesterreich F, Herzel L, Straube K, Hujer K, Howard J, Neugebauer KM. 2016. Splicing of Nascent RNA Coincides with Intron Exit from RNA Polymerase II. *Cell* **165**:372–381.
Carrillo Oesterreich F, Preibisch S, Neugebauer KM. 2010. Global analysis of nascent RNA reveals transcriptional pausing in terminal exons. *Mol Cell* **40**:571–581.
Castro-Viñuelas R, Sanjurjo-Rodríguez C, Piñeiro-Ramil M, Hermida-Gómez T, Fuentes-Boquete IM, de Toro-Santos FJ, Blanco-García FJ, Díaz-Prado SM. 2018. Induced pluripotent stem cells for cartilage repair: current status and future perspectives. *Eur Cell Mater* **36**:96–109.
Caterson B, Flannery CR, Hughes CE, Little CB. 2000. Mechanisms involved in cartilage proteoglycan catabolism. *Matrix Biol* **19**:333–344.
Chan CKF, Seo EY, Chen JY, Lo D, McArdle A, Sinha R, Tevlin R, Seita J, Vincent-Tompkins J, Wearda T, Lu W-J, Senarath-Yapa K, Chung MT, Marecic O, Tran M, Yan KS, Upton R, Walmsley GG, Lee AS, Sahoo D, Kuo CJ, Weissman IL, Longaker MT. 2015. Identification and specification of the mouse skeletal stem cell. *Cell* **160**:285–298.
Chen F, Tillberg PW, Boyden ES. 2015. Optical imaging. Expansion microscopy. *Science* **347**:543–548.
Chen F, Wassie AT, Cote AJ, Sinha A, Alon S, Asano S, Daugharthy ER, Chang J-B, Marblestone A, Church GM, Raj A, Boyden ES. 2016. Nanoscale imaging of RNA with expansion microscopy. *Nat Methods* **13**:679–684.
Cheng T, Maddox NC, Wong AW, Rahnama R, Kuo AC. 2012. Comparison of gene expression patterns in articular cartilage and dedifferentiated articular chondrocytes. *J Orthop Res* **30**:234–245.
Chen Y, Zhang Y, Wang Y, Zhang L, Brinkman EK, Adam SA, Goldman R, van Steensel B, Ma J, Belmont AS. 2018. Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *J Cell Biol* **217**:4025–4048.
Choi PJ, Cai L, Frieda K, Xie XS. 2008. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science* **322**:442–446.
Chubb JR, Trcek T, Shenoy SM, Singer RH. 2006. Transcriptional pulsing of a developmental gene. *Curr Biol* **16**:1018–1025.
Chubinskaya S, Haudenschild D, Gasser S, Stannard J, Krettek C, Borrelli J Jr. 2015. Articular Cartilage Injury and Potential Remedies. *J Orthop Trauma* **29 Suppl 12**:S47–52.
Churchman LS, Weissman JS. 2011. Nascent transcript sequencing visualizes transcription at nucleotide resolution. *Nature* **469**:368–373.
Coleman CM, Flug JA, Major N. 2017. Imaging of Cartilage in the Athlete. *Clin Sports Med* **36**:427–445.

Colter DC, Sekiya I, Prockop DJ. 2001. Identification of a subpopulation of rapidly self-renewing and multipotential adult stem cells in colonies of human marrow stromal cells. *Proc Natl Acad Sci U S A* **98**:7841–7845.
Cote AJ, McLeod CM, Farrell MJ, McClanahan PD, Dunagin MC, Raj A, Mauck RL. 2016. Single-cell differences in matrix gene expression do not predict matrix deposition. *Nat Commun* **7**:10865.
Coulon A, Ferguson ML, de Turris V, Palangat M, Chow CC, Larson DR. 2014. Kinetic competition during the transcription cycle results in stochastic RNA processing. *Elife* **3**. doi:10.7554/eLife.03939
Crosetto N, Bienko M, van Oudenaarden A. 2015. Spatially resolved transcriptomics and beyond. *Nat Rev Genet* **16**:57–66.
Curtis KM, Gomez LA, Rios C, Garbayo E, Raval AP, Perez-Pinzon MA, Schiller PC. 2010. EF1alpha and RPL13a represent normalization genes suitable for RT-qPCR analysis of bone marrow derived mesenchymal stem cells. *BMC Mol Biol* **11**:61.
Custódio N, Carmo-Fonseca M. 2016. Co-transcriptional splicing and the CTD code. *Crit Rev Biochem Mol Biol* **51**:395–411.
Darling EM, Athanasiou KA. 2005. Rapid phenotypic changes in passaged articular chondrocyte subpopulations. *J Orthop Res* **23**:425–432.
Dar RD, Razooky BS, Singh A, Trimeloni TV, McCollum JM, Cox CD, Simpson ML, Weinberger LS. 2012. Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proc Natl Acad Sci U S A* **109**:17454–17459.
Davis-Turak JC, Allison K, Shokhirev MN, Ponomarenko P, Tsimring LS, Glass CK, Johnson TL, Hoffmann A. 2015. Considering the kinetics of mRNA synthesis in the analysis of the genome and epigenome reveals determinants of co-transcriptional splicing. *Nucleic Acids Res* **43**:699–707.
de la Mata M, Alonso CR, Kadener S, Fededa JP, Blaustein M, Pelisch F, Cramer P, Bentley D, Kornblihtt AR. 2003. A slow RNA polymerase II affects alternative splicing in vivo. *Mol Cell* **12**:525–532.
de la Mata M, Lafaille C, Kornblihtt AR. 2010. First come, first served revisited: factors affecting the same alternative splicing event have different effects on the relative rates of intron removal. *RNA* **16**:904–912.
DeLise AM, Fischer L, Tuan RS. 2000. Cellular interactions and signaling in cartilage development. *Osteoarthritis Cartilage* **8**:309–334.
Delorme B, Ringe J, Pontikoglou C, Gaillard J, Langonné A, Sensebé L, Noël D, Jorgensen C, Häupl T, Charbord P. 2009. Specific Lineage-Priming of Bone Marrow Mesenchymal Stem Cells Provides the Molecular Framework for Their Plasticity: Differentiation Potential of Human BM MSCs. *Stem Cells* **27**:1142–1151.
D'Ippolito G, Schiller PC, Ricordi C, Roos BA, Howard GA. 1999. Age-related osteogenic potential of mesenchymal stromal stem cells from human vertebral bone marrow. *J Bone Miner Res* **14**:1115–1122.
Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15–21.
Dölken, L., Ruzsics, Z., Rädle, B., Friedel, C.C., Zimmer, R., Mages, J., Hoffmann, R., Dickinson, P., Forster, T., Ghazal, P., et al. (2008). High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. RNA *14*, 1959–1972.
Dominici M, Le Blanc K, Mueller I, Slaper-Cortenbach I, Marini F, Krause D, Deans R,

Keating A, Prockop D, Horwitz E. 2006. Minimal criteria for defining multipotent mesenchymal stromal cells. The International Society for Cellular Therapy position statement. *Cytotherapy* **8**:315–317.
Drexler HL, Choquet K, Stirling Churchman L. 2019. Human co-transcriptional splicing kinetics and coordination revealed by direct nascent RNA sequencing. *bioRxiv*. doi:10.1101/611020
Drury JL, Mooney DJ. 2003. Hydrogels for tissue engineering: scaffold design variables and applications. *Biomaterials* **24**:4337–4351.
Eldar A, Elowitz MB. 2010. Functional roles for noise in genetic circuits. *Nature* **467**:167–173.
Elima K, Vuorio E. 1989. Expression of mRNAs for collagens and other matrix components in dedifferentiating and redifferentiating human chondrocytes in culture. *FEBS Lett* **258**:195–198.
Eser P, Wachutka L, Maier KC, Demel C, Boroni M, Iyer S, Cramer P, Gagneur J. 2016. Determinants of RNA metabolism in the Schizosaccharomyces pombe genome. *Mol Syst Biol* **12**:857.
Farrell MJ, Shin JI, Smith LJ, Mauck RL. 2015. Functional consequences of glucose and oxygen deprivation on engineered mesenchymal stem cell-based cartilage constructs. *Osteoarthritis Cartilage* **23**:134–142.
Faustino NA, Cooper TA. 2003. Pre-mRNA splicing and human disease. *Genes Dev* **17**:419–437.
Femino AM, Fay FS, Fogarty K, Singer RH. 1998. Visualization of single RNA transcripts in situ. *Science* **280**:585–590.
Frank O, Heim M, Jakob M, Barbero A, Schäfer D, Bendik I, Dick W, Heberer M, Martin I. 2002. Real-time quantitative RT-PCR analysis of human bone marrow stromal cells during osteogenic differentiation in vitro. *J Cell Biochem* **85**:737–746.
Freeman BT, Jung JP, Ogle BM. 2015. Single-Cell RNA-Seq of Bone Marrow-Derived Mesenchymal Stem Cells Reveals Unique Profiles of Lineage Priming. *PLoS One* **10**. doi:10.1371/journal.pone.0136199
Gandhi SJ, Zenklusen D, Lionnet T, Singer RH. 2011. Transcription of functionally related constitutive genes is not coordinated. *Nat Struct Mol Biol* **18**:27–34.
Gazzoli I, Pulyakhina I, Verwey NE, Ariyurek Y, Laros JFJ, 't Hoen PAC, Aartsma-Rus A. 2016. Non-sequential and multi-step splicing of the dystrophin transcript. *RNA Biol* **13**:290–305.
Glowacki J, Trepman E, Folkman J. 1983. Cell shape and phenotypic expression in chondrocytes. *Proc Soc Exp Biol Med* **172**:93–98.
Golding I, Paulsson J, Zawilski SM, Cox EC. 2005. Real-time kinetics of gene activity in individual bacteria. *Cell* **123**:1025–1036.
Goldstrohm AC, Greenleaf AL, Garcia-Blanco MA. 2001. Co-transcriptional splicing of pre-messenger RNAs: considerations for the mechanism of alternative splicing. *Gene* **277**:31–47.
González-Cruz RD, Fonseca VC, Darling EM. 2012. Cellular mechanical properties reflect the differentiation potential of adipose-derived mesenchymal stem cells. *Proc Natl Acad Sci U S A* **109**:E1523–9.
Gruber AJ, Schmidt R, Gruber AR, Martin G, Ghosh S, Belmadani M, Keller W, Zavolan M. 2016. A comprehensive analysis of 3' end sequencing data sets reveals novel polyadenylation signals and the repressive role of heterogeneous ribonucleoprotein C on cleavage and polyadenylation. *Genome Res* **26**:1145–1159.

Harlen KM, Churchman LS. 2017. Subgenic Pol II interactomes identify region-specific transcription elongation regulators. *Mol Syst Biol* **13**:900.
Harlen KM, Trotta KL, Smith EE, Mosaheb MM, Fuchs SM, Churchman LS. 2016. Comprehensive RNA Polymerase II Interactomes Reveal Distinct and Varied Roles for Each Phospho-CTD Residue. *Cell Rep* **15**:2147–2158.
Herzel L, Ottoz DSM, Alpert T, Neugebauer KM. 2017. Splicing and transcription touch base: co-transcriptional spliceosome assembly and function. *Nat Rev Mol Cell Biol* **18**:637–650.
Huang AH, Farrell MJ, Mauck RL. 2010. Mechanics and mechanobiology of mesenchymal stem cell-based engineered cartilage. *J Biomech* **43**:128–136.
Huang BJ, Hu JC, Athanasiou KA. 2016. Cell-based tissue engineering strategies used in the clinical repair of articular cartilage. *Biomaterials* **98**:1–22.
Huranová M, Ivani I, Benda A, Poser I, Brody Y, Hof M, Shav-Tal Y, Neugebauer KM, Stanek D. 2010. The differential interaction of snRNPs with pre-mRNA reveals splicing kinetics in living cells. *J Cell Biol* **191**:75–86.
Hwang B, Lee JH, Bang D. 2018. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med* **50**:96.
Ingolia NT. 2016. Ribosome Footprint Profiling of Translation throughout the Genome. *Cell* **165**:22–33.
Itzkovitz S, van Oudenaarden A. 2011. Validating transcripts with probes and imaging technology. *Nat Methods* **8**:S12–9.
Junker JP, van Oudenaarden A. 2014. Every cell is special: genome-wide studies add a new dimension to single-cell biology. *Cell* **157**:8–11.
Kapp LD, Lorsch JR. 2004. The molecular mechanics of eukaryotic translation. *Annu Rev Biochem* **73**:657–704.
Keren H, Lev-Maor G, Ast G. 2010. Alternative splicing and evolution: diversification, exon definition and function. *Nat Rev Genet* **11**:345–355.
Kessler O, Jiang Y, Chasin LA. 1993. Order of intron removal during splicing of endogenous adenine phosphoribosyltransferase and dihydrofolate reductase pre-mRNA. *Mol Cell Biol* **13**:6211–6222.
Kim S, Park C, Jun Y, Lee S, Jung Y, Kim J. 2018. Integrative Profiling of Alternative Splicing Induced by U2AF1 S34F Mutation in Lung Adenocarcinoma Reveals a Mechanistic Link to Mitotic Stress. *Mol Cells* **41**:733–741.
Kim SW, Taggart AJ, Heintzelman C, Cygan KJ, Hull CG, Wang J, Shrestha B, Fairbrother WG. 2017. Widespread intra-dependencies in the removal of introns from human transcripts. *Nucleic Acids Res* **45**:9503–9513.
King FW, Liszewski W, Ritner C, Bernstein HS. 2011. High-throughput tracking of pluripotent human embryonic stem cells with dual fluorescence resonance energy transfer molecular beacons. *Stem Cells Dev* **20**:475–484.
Knudson CB. 1993. Hyaluronan receptor-directed assembly of chondrocyte pericellular matrix. *J Cell Biol* **120**:825–834.
Kumar D, Lassar AB. 2009. The transcriptional activity of Sox9 in chondrocytes is regulated by RhoA signaling and actin polymerization. *Mol Cell Biol* **29**:4262–4273.
Kumar RM, Cahan P, Shalek AK, Satija R, DaleyKeyser AJ, Li H, Zhang J, Pardee K, Gennert D, Trombetta JJ, Ferrante TC, Regev A, Daley GQ, Collins JJ. 2014. Deconstructing transcriptional heterogeneity in pluripotent stem cells. *Nature* **516**:56–61.
Lahm H, Doppler S, Dreßen M, Werner A, Adamczyk K, Schrambke D, Brade T,

Laugwitz K-L, Deutsch M-A, Schiemann M, Lange R, Moretti A, Krane M. 2015. Live fluorescent RNA-based detection of pluripotency gene expression in embryonic and induced pluripotent stem cells of different species. *Stem Cells* **33**:392–402.
Larsson HM, Lee ST, Roccio M, Velluto D, Lutolf MP, Frey P, Hubbell JA. 2012. Sorting live stem cells based on Sox2 mRNA expression. *PLoS One* **7**:e49874.
Levesque MJ, Raj A. 2013. Single-chromosome transcriptional profiling reveals chromosomal gene expression regulation. *Nat Methods* **10**:246–248.
Liao J, Qu Y, Chu B, Zhang X, Qian Z. 2015. Biodegradable CSMA/PECA/Graphene Porous Hybrid Scaffold for Cartilage Tissue Engineering. *Sci Rep* **5**:9879.
Lin Z, Fitzgerald JB, Xu J, Willers C, Wood D, Grodzinsky AJ, Zheng MH. 2008. Gene expression profiles of human chondrocytes during passaged monolayer cultivation. *J Orthop Res* **26**:1230–1237.
Lugowski A, Nicholson B, Rissland OS. 2018. Determining mRNA half-lives on a transcriptome-wide scale. *Methods* **137**:90–98.
Lukacsovich D, Winterer J, Que L, Luo W, Lukacsovich T, Földy C. 2019. Single-Cell RNA-Seq Reveals Developmental Origins and Ontogenetic Stability of Neurexin Alternative Splicing Profiles. *Cell Rep* **27**:3752–3759.e4.
Luo W, Guo C, Zheng J, Chen TL, Wang PY, Vertel BM, Tanzer ML. 2000. Aggrecan from start to finish. *J Bone Miner Metab* **18**:51–56.
Maamar H, Raj A, Dubnau D. 2007. Noise in gene expression determines cell fate in Bacillus subtilis. *Science* **317**:526–529.
Ma B, Leijten JCH, Wu L, Kip M, van Blitterswijk CA, Post JN, Karperien M. 2013. Gene expression profiling of dedifferentiated human articular chondrocytes in monolayer culture. *Osteoarthritis Cartilage* **21**:599–603.
Mackay AM, Beck SC, Murphy JM, Barry FP, Chichester CO, Pittenger MF. 1998. Chondrogenic differentiation of cultured human mesenchymal stem cells from marrow. *Tissue Eng* **4**:415–428.
Marble HD, Sutermaster BA, Kanthilal M, Fonseca VC, Darling EM. 2014. Gene expression-based enrichment of live cells from adipose tissue produces subpopulations with improved osteogenic potential. *Stem Cell Res Ther* **5**:145.
Mareddy S, Crawford R, Brooke G, Xiao Y. 2007. Clonal isolation and characterization of bone marrow stromal cells from patients with osteoarthritis. *Tissue Eng* **13**:819–829.
Maroudas A, Wachtel E, Grushko G, Katz EP, Weinberg P. 1991. The effect of osmotic and mechanical pressures on water partitioning in articular cartilage. *Biochim Biophys Acta* **1073**:285–294.
Martin RM, Rino J, Carvalho C, Kirchhausen T, Carmo-Fonseca M. 2013. Live-cell visualization of pre-mRNA splicing with single-molecule sensitivity. *Cell Rep* **4**:1144–1155.
Mauck RL, Soltz MA, Wang CC, Wong DD, Chao PH, Valhmu WB, Hung CT, Ateshian GA. 2000. Functional tissue engineering of articular cartilage through dynamic loading of chondrocyte-seeded agarose gels. *J Biomech Eng* **122**:252–260.
Mauck R, Yuan X, Tuan R. 2006. Chondrogenic differentiation and functional maturation of bovine mesenchymal stem cells in long-term agarose culture. *Osteoarthritis Cartilage* **14**:179–189.
Mayer A, Churchman LS. 2017. A Detailed Protocol for Subcellular RNA Sequencing (subRNA-seq). *Curr Protoc Mol Biol* **120**:4.29.1–4.29.18.
Mayer A, di Iulio J, Maleri S, Eser U, Vierstra J, Reynolds A, Sandstrom R, Stamatoyannopoulos JA, Churchman LS. 2015. Native elongating transcript

sequencing reveals human transcriptional activity at nucleotide resolution. *Cell* **161**:541–554.
McLeod CM, Mauck RL. 2017. On the origin and impact of mesenchymal stem cell heterogeneity: new insights and emerging tools for single cell analysis. *Eur Cell Mater* **34**:217–231.
Milligan L, Huynh-Thu VA, Delan-Forino C, Tuck A, Petfalski E, Lombraña R, Sanguinetti G, Kudla G, Tollervey D. 2016. Strand-specific, high-resolution mapping of modified RNA polymerase II. *Mol Syst Biol* **12**:874.
Miura Y, Fitzsimmons JS, Commisso CN, Gallay SH, O'Driscoll SW. 1994. Enhancement of periosteal chondrogenesis in vitro. Dose-response for transforming growth factor-beta 1 (TGF-beta 1). *Clin Orthop Relat Res* 271–280.
Molina N, Suter DM, Cannavo R, Zoller B, Gotic I, Naef F. 2013. Stimulus-induced modulation of transcriptional bursting in a single mammalian gene. *Proc Natl Acad Sci U S A* **110**:20563–20568.
Morlando M, Ballarino M, Gromak N, Pagano F, Bozzoni I, Proudfoot NJ. 2008. Primary microRNA transcripts are processed co-transcriptionally. *Nat Struct Mol Biol* **15**:902–909.
Motorin Y, Helm M. 2019. Methods for RNA Modification Mapping Using Deep Sequencing: Established and New Emerging Technologies. *Genes* **10**. doi:10.3390/genes10010035
Muraglia A, Cancedda R, Quarto R. 2000. Clonal mesenchymal progenitors from human bone marrow differentiate in vitro according to a hierarchical model. *J Cell Sci* **113 (Pt 7)**:1161–1166.
Ng J, Bernhard J, Vunjak-Novakovic G. 2016. Mesenchymal Stem Cells for Osteochondral Tissue Engineering In: Gnecchi M, editor. Mesenchymal Stem Cells: Methods and Protocols. New York, NY: Springer New York. pp. 35–54.
Nicholls C, Li H, Liu J-P. 2012. GAPDH: a common enzyme with uncommon functions. *Clin Exp Pharmacol Physiol* **39**:674–679.
Nimmo RA, May GE, Enver T. 2015. Primed and ready: understanding lineage commitment through single cell analysis. *Trends Cell Biol* **25**:459–467.
Nojima T, Gomes T, Grosso ARF, Kimura H, Dye MJ, Dhir S, Carmo-Fonseca M, Proudfoot NJ. 2015. Mammalian NET-Seq Reveals Genome-wide Nascent Transcription Coupled to RNA Processing. *Cell* **161**:526–540.
Nojima T, Rebelo K, Gomes T, Grosso AR, Proudfoot NJ, Carmo-Fonseca M. 2018. RNA Polymerase II Phosphorylated on CTD Serine 5 Interacts with the Spliceosome during Co-transcriptional Splicing. *Mol Cell* **72**:369–379.e4.
Octavio LM, Gedeon K, Maheshri N. 2009. Epigenetic and conventional regulation is distributed among activators of FLO11 allowing tuning of population-level heterogeneity in its expression. *PLoS Genet* **5**:e1000673.
Padovan-Merhar O, Nair GP, Biaesch AG, Mayer A, Scarfone S, Foley SW, Wu AR, Churchman LS, Singh A, Raj A. 2015. Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. *Mol Cell* **58**:339–352.
Palangat M, Anastasakis DG, Fei DL, Lindblad KE, Bradley R, Hourigan CS, Hafner M, Larson DR. 2019. The splicing factor U2AF1 contributes to cancer progression through a noncanonical role in translation regulation. *Genes Dev* **33**:482–497.
Panadero JA, Lanceros-Mendez S, Ribelles JLG. 2016. Differentiation of mesenchymal stem cells for cartilage tissue engineering: Individual and synergetic effects of three-

dimensional environment and mechanical loading. *Acta Biomater* **33**:1–12.
Pandya-Jones A, Bhatt DM, Lin C-H, Tong A-J, Smale ST, Black DL. 2013. Splicing kinetics and transcript release from the chromatin compartment limit the rate of Lipid A-induced gene expression. *RNA* **19**:811–827.
Pandya-Jones A, Black DL. 2009. Co-transcriptional splicing of constitutive and alternative exons. *RNA* **15**:1896–1908.
Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40**:1413–1415.
Pelttari K, Steck E, Richter W. 2008. The use of mesenchymal stem cells for chondrogenesis. *Injury* **39 Suppl 1**:S58–65.
Phinney DG. 2012. Functional heterogeneity of mesenchymal stem cells: implications for cell therapy. *J Cell Biochem* **113**:2806–2812.
Phinney DG, Kopen G, Righter W, Webster S, Tremain N, Prockop DJ. 1999. Donor variation in the growth properties and osteogenic potential of human marrow stromal cells. *J Cell Biochem* **75**:424–436.
Pittenger MF, Mackay AM, Beck SC, Jaiswal RK, Douglas R, Mosca JD, Moorman MA, Simonetti DW, Craig S, Marshak DR. 1999. Multilineage potential of adult human mesenchymal stem cells. *Science* **284**:143–147.
Ponce ML, Koelling S, Kluever A, Heinemann DEH, Miosge N, Wulf G, Frosch K-H, Schütze N, Hufner M, Siggelkow H. 2008. Coexpression of osteogenic and adipogenic differentiation markers in selected subpopulations of primary human mesenchymal progenitor cells. *J Cell Biochem* **104**:1342–1355.
Puetzer JL, Petitte JN, Loboa EG. 2010. Comparative review of growth factors for induction of three-dimensional in vitro chondrogenesis in human mesenchymal stem cells isolated from bone marrow and adipose tissue. *Tissue Eng Part B Rev* **16**:435–444.
Rabani M, Levin JZ, Fan L, Adiconis X, Raychowdhury R, Garber M, Gnirke A, Nusbaum C, Hacohen N, Friedman N, Amit I, Regev A. 2011. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat Biotechnol* **29**:436–442.
Rabani M, Raychowdhury R, Jovanovic M, Rooney M, Stumpo DJ, Pauli A, Hacohen N, Schier AF, Blackshear PJ, Friedman N, Amit I, Regev A. 2014. High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. *Cell* **159**:1698–1710.
Rai V, Dilisio MF, Dietz NE, Agrawal DK. 2017. Recent strategies in cartilage repair: A systemic review of the scaffold development and tissue engineering. *J Biomed Mater Res A* **105**:2343–2354.
Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. 2006. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol* **4**:e309.
Raj A, Rifkin SA, Andersen E, van Oudenaarden A. 2010. Variability in gene expression underlies incomplete penetrance. *Nature* **463**:913–918.
Raj A, van den Bogaard P, Rifkin SA, van Oudenaarden A, Tyagi S. 2008. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat Methods* **5**:877–879.
Raj A, van Oudenaarden A. 2009. Single-molecule approaches to stochastic gene expression. *Annu Rev Biophys* **38**:255–270.
Raj A, van Oudenaarden A. 2008. Nature, nurture, or chance: stochastic gene

expression and its consequences. *Cell* **135**:216–226.
Ramaswami G, Zhang R, Piskol R, Keegan LP, Deng P, O'Connell MA, Li JB. 2013. Identifying RNA editing sites using RNA sequencing data alone. *Nat Methods* **10**:128–132.
Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, Müller M. 2011. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12**:77.
Russell KC, Phinney DG, Lacey MR, Barrilleaux BL, Meyertholen KE, O'Connor KC. 2010. In vitro high-capacity assay to quantify the clonal heterogeneity in trilineage potential of mesenchymal stem cells reveals a complex hierarchy of lineage commitment. *Stem Cells* **28**:788–798.
Sakata R, Iwakura T, Reddi AH. 2015. Regeneration of Articular Cartilage Surface: Morphogens, Cells, and Extracellular Matrix Scaffolds. *Tissue Eng Part B Rev* **21**:461–473.
Saldi T, Cortazar MA, Sheridan RM, Bentley DL. 2016. Coupling of RNA Polymerase II Transcription Elongation with Pre-mRNA Splicing. *J Mol Biol* **428**:2623–2635.
Schiltz JR, Mayne R, Holtzer H. 1973. The Synthesis of Collagen and Glycosaminoglycans by Dedifferentiated Chondroblasts in Culture. *Differentiation* **1**:97–108.
Schwalb, B., Michel, M., Zacher, B., Frühauf, K., Demel, C., Tresch, A., Gagneur, J., and Cramer, P. (2016). TT-seq maps the human transient transcriptome. Science *352*, 1225–1228.
Schwarz S, Koerber L, Elsaesser AF, Goldberg-Bockhorn E, Seitz AM, Dürselen L, Ignatius A, Walther P, Breiter R, Rotter N. 2012. Decellularized cartilage matrix as a novel biomatrix for cartilage tissue-engineering applications. *Tissue Eng Part A* **18**:2195–2209.
Scuteri A, Donzelli E, Foudah D, Caldara C, Redondo J, D'Amico G, Tredici G, Miloso M. 2014. Mesengenic differentiation: comparison of human and rat bone marrow mesenchymal stem cells. *Int J Stem Cells* **7**:127–134.
Sekiya I, Larson BL, Vuoristo JT, Cui J-G, Prockop DJ. 2003. Adipogenic Differentiation of Human Adult Stem Cells From Bone Marrow Stroma (MSCs). *J Bone Miner Res* **19**:256–264.
Shaffer SM, Dunagin MC, Torborg SR, Torre EA, Emert B, Krepler C, Beqiri M, Sproesser K, Brafford PA, Xiao M, Eggan E, Anastopoulos IN, Vargas-Garcia CA, Singh A, Nathanson KL, Herlyn M, Raj A. 2017. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nature* **546**:431–435.
Shah K, Tyagi S. 2013. Barriers to transmission of transcriptional noise in a c-fos c-jun pathway. *Mol Syst Biol* **9**:687.
Singh J, Padgett RA. 2009. Rates of in situ transcription and splicing in large human genes. *Nat Struct Mol Biol* **16**:1128–1133.
Song L. 2004. Transdifferentiation potential of human mesenchymal stem cells derived from bone marrow. *The FASEB Journal*.
Sophia Fox AJ, Bedi A, Rodeo SA. 2009. The basic science of articular cartilage: structure, composition, and function. *Sports Health* **1**:461–468.
Spector DL, Lamond AI. 2011. Nuclear speckles. *Cold Spring Harb Perspect Biol* **3**. doi:10.1101/cshperspect.a000646
Stewart M. 2019. Polyadenylation and nuclear export of mRNAs. *J Biol Chem* **294**:2977–2987.

Stockwell RA. 1971. The interrelationship of cell density and cartilage thickness in mammalian articular cartilage. *J Anat* **109**:411–421.
Stokes DG, Liu G, Dharmavaram R, Hawkins D, Piera-Velazquez S, Jimenez SA. 2001. Regulation of type-II collagen gene expression during human chondrocyte de-differentiation and recovery of chondrocyte-specific phenotype in culture involves Sry-type high-mobility-group box (SOX) transcription factors. *Biochem J* **360**:461–470.
Stuart T, Satija R. 2019. Integrative single-cell analysis. *Nat Rev Genet* **20**:257–272.
Süel GM, Kulkarni RP, Dworkin J, Garcia-Ojalvo J, Elowitz MB. 2007. Tunability and noise dependence in differentiation dynamics. *Science* **315**:1716–1719.
Suter DM, Molina N, Gatfield D, Schneider K, Schibler U, Naef F. 2011. Mammalian genes are transcribed with widely different bursting kinetics. *Science* **332**:472–474.
Symmons O, Raj A. 2016. What's Luck Got to Do with It: Single Cells, Multiple Fates, and Biological Nondeterminism. *Mol Cell* **62**:788–802.
Takahashi H, Lassmann T, Murata M, Carninci P. 2012. 5' end-centered expression profiling using cap-analysis gene expression and next-generation sequencing. *Nat Protoc* **7**:542–561.
Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, Lao K, Surani MA. 2009. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* **6**:377–382.
Tilgner H, Knowles DG, Johnson R, Davis CA, Chakrabortty S, Djebali S, Curado J, Snyder M, Gingeras TR, Guigó R. 2012. Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res* **22**:1616–1625.
Toh WS, Cao T. 2014. Derivation of Chondrogenic Cells from Human Embryonic Stem Cells for Cartilage Tissue Engineering. *Methods Mol Biol*. doi:10.1007/7651_2014_89
Toh WS, Lee EH, Cao T. 2011. Potential of human embryonic stem cells in cartilage tissue engineering and regenerative medicine. *Stem Cell Rev Rep* **7**:544–559.
Tsai MJ, Ting AC, Nordstrom JL, Zimmer W, O'Malley BW. 1980. Processing of high molecular weight ovalbumin and ovomucoid precursor RNAs to messenger RNA. *Cell* **22**:219–230.
Valencia P, Dias AP, Reed R. 2008. Splicing promotes rapid and efficient mRNA export in mammalian cells. *Proc Natl Acad Sci U S A* **105**:3386–3391.
Vargas DY, Shah K, Batish M, Levandoski M, Sinha S, Marras SAE, Schedl P, Tyagi S. 2011. Single-molecule imaging of transcriptionally coupled and uncoupled splicing. *Cell* **147**:1054–1065.
Vertel BM. 1995. The ins and outs of aggrecan. *Trends Cell Biol* **5**:458–464.
Waks Z, Klein AM, Silver PA. 2011. Cell-to-cell variability of alternative RNA splicing. *Mol Syst Biol* **7**:506.
Wallace EWJ, Beggs JD. 2017. Extremely fast and incredibly close: cotranscriptional splicing in budding yeast. *RNA* **23**:601–610.
Wang W, Rigueur D, Lyons KM. 2014. TGFβ signaling in cartilage development and maintenance. *Birth Defects Res C Embryo Today* **102**:37–51.
Wang Y, Liu J, Huang BO, Xu Y-M, Li J, Huang L-F, Lin J, Zhang J, Min Q-H, Yang W-M, Wang X-Z. 2015. Mechanism of alternative splicing and its regulation. *Biomed Rep* **3**:152–158.
Wernet MF, Mazzoni EO, Celik A, Duncan DM, Duncan I, Desplan C. 2006. Stochastic

spineless expression creates the retinal mosaic for colour vision. *Nature* **440**:174–180.
Wilkinson ME, Lin P-C, Plaschka C, Nagai K. 2018. Cryo-EM Studies of Pre-mRNA Splicing: From Sample Preparation to Model Visualization. *Annu Rev Biophys* **47**:175–199.
Windhager L, Bonfert T, Burger K, Ruzsics Z, Krebs S, Kaufmann S, Malterer G, L'Hernault A, Schilhabel M, Schreiber S, Rosenstiel P, Zimmer R, Eick D, Friedel CC, Dölken L. 2012. Ultrashort and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution. *Genome Res* **22**:2031–2042.
Woo SL, Buckwalter JA. 1988. AAOS/NIH/ORS workshop. Injury and repair of the musculoskeletal soft tissues. Savannah, Georgia, June 18-20, 1987. *J Orthop Res* **6**:907–931.
Wuarin J, Schibler U. 1994. Physical isolation of nascent RNA chains transcribed by RNA polymerase II: evidence for cotranscriptional splicing. *Mol Cell Biol* **14**:7219–7225.
Yang M, Wu J, Wu S-H, Bi A-D, Liao DJ. 2012. Splicing of mouse p53 pre-mRNA does not always follow the "first come, first served" principle and may be influenced by cisplatin treatment and serum starvation. *Mol Biol Rep* **39**:9247–9256.
Yoshimoto H, Shin YM, Terai H, Vacanti JP. 2003. A biodegradable nanofiber scaffold by electrospinning and its potential for bone tissue engineering. *Biomaterials* **24**:2077–2082.
Zenklusen D, Larson DR, Singer RH. 2008. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat Struct Mol Biol* **15**:1263–1271.
Zhang G, Taneja KL, Singer RH, Green MR. 1994. Localization of pre-mRNA splicing in mammalian nuclei. *Nature* **372**:809–812.
Zopf CJ, Quinn K, Zeidman J, Maheshri N. 2013. Cell-cycle dependence of transcription dominates noise in gene expression. *PLoS Comput Biol* **9**:e1003161.